공학석사 학위논문

베이지안 신경망 냉동기 모델의 인식론적 및 내재적 불확실성 분석

Epistemic and Aleatoric Uncertainty of Bayesian Neural Network Model for a Chiller

2020년 8월

서울대학교 대학원

건축학과 건축공학전공

김 재 민

베이지안 신경망 냉동기 모델의 인식론적 및 내재적 불확실성 분석

Epistemic and Aleatoric Uncertainty of Bayesian Neural Network Model for a Chiller

지도교수 박 철 수

이 논문을 공학석사 학위논문으로 제출함

2020년 8월

서울대학교 대학원

건축학과 건축공학전공

김 재 민

김재민의 석사 학위논문을 인준함

2020년 8월

위 원 장 여 명 석 (인)

부위원장 박 철 수 (인)

위 원 김 선 숙 (인)

국문초록

기계학습 모델은 입력 및 출력데이터 간의 관계를 통계적으로 학습하는 모델이기 때문에, 데이터만 주어진다면 수학적 모델을 작성하지 않고 시뮬레이션 모델을 제작할 수 있다. 이러한 장점으로, 기계학습은 건물에너지 시뮬레이션 분야에서 수요예측, 최적제어, 고장진단 등 다양한 주제로 연구되고 있다. 하지만, 기계학습 모델은 black-box model로서의 한계로 인하여 입력변수와 출력변수 간의 인과관계를 설명할 수 없으며, 모델에 내재한 불확실성이 존재한다.

기계학습 모델의 불확실성은 모델이 예측한 결과를 신뢰할 수 있는 정도를 의미하며, 불확실성의 발생원인에 따라 인식론적(epistemic) 불확실성과 내재적(aleatoric) 불확실성으로 구분할 수 있다. 인식론적 불확실성은 데이터 또는 지식의 부재(lack of data or knowledge)로 인해 발생하는 불확실성으로, 신경망의 가중치에 의한 불확실성을 의미한다. 내재적 불확실성은 자연 현상의 본질적인(intrinsic) 무작위성(randomness)에 의해 발생하는 불확실성으로, 자연 현상의 무작위성이란 계측 데이터의 센싱 오차 또는 계측 시스템 자체의 오작동 등에 의해 발생하는 노이즈 및 이상치를 의미한다.

기계학습 모델의 불확실성에 대한 정량적 평가는 모델 예측 결과에 대한 신뢰도 정보를 제공함으로써 모델 성능에 대한 평가 지표를 제공해준다. 예를 들어 불확실성에 대한 평가 없이 단순히 검증 기간 데이터에 대한 예측 오차만으로 모델의 성능을 평가하고 사용한다면, 하이퍼 파라미터 선택, 초깃값 설정, 최적화 과정에서의 확률적 특성 등에 의해 매번 다른 가중치를 학습할 수 있어 안정적인 예측 성능을 확보하기 어려울 수 있다. 또한, 인식론적 및 내재적 불확실성을 분리함으로써 훈련데이터의 양적, 질적 품질을 평가하는 지표로써 사용될 수 있으며, 최적제어의 목적함수 또는 의사결정의 수단으로 사용하는 등 분석자에게 다양한 정보를 제공할 수 있다.

하지만, 기존에 사용되는 대부분의 기계학습 알고리즘들(인공신경망, 서포트 벡터 머신 등)은 모델의 불확실성을 평가하기 어렵다. 베이지안 신경망(Bayesian Neural Network, BNN)은 모델의 가중치에 대해 확률분포의 형태로 추정함으로써, 신경망의 확률적 해석을 가능하게 한 알고리즘으로, 신경망의 인식론적 및 내재적 불확실성을 분리하여 정량화할 수 있는 장점이 있다. BNN의 이론적 기반은 매우 간단하지만, 연산량의 한계로 인하여 실질적인 구현이 어렵다는 단점이 있다. BNN의 실용적 구현을 위해 다양한 연구가 진행되고 있으며, 본 논문에서는 Yarin Gal(2016)이 제안한 몬테카를로 드랍아웃(Monte Carlo Dropout) 방법을 이용하여 BNN을 구현하였다.

본 연구에서는 서울시 소재의 실제 업무용 건물에 설치된 BEMS 데이터를 사용하여 BNN 모델을 제작하고, 모델에 내재된 인식론적 및 내재적 불확실성을 분리하여 정량화하였다. BNN 모델은 냉동기 가동 조건(입수온도, 유량, 전력)에 따라 변화하는 냉동기의 COP를 예측하는 모델로 제작되었으며, 훈련데이터의 기간과 데이터 내 이상치 처리 여부에 따라 BNN 모델을 구분하여 각각의 모델을 예측 오차, 인식론적 불확실성, 내재적 불확실성의 측면에서 비교분석 하였다. 분석 결과, 제작된 4개의 BNN 모델 모두 검증 기간에 대한 예측 성능은 우수했지만, 내재된 불확실성의 크기는 모두 달랐다. 정격 COP 4.81에 대한 불확실성 크기의 비율이 2.8%에서 11.6%까지 다양했으며, 불확실성이 클수록 모델 예측 결과의 무작위성 또한 증가하였다. 또한, 각 모델의 불확실성을 인식론적, 내재적 불확실성으로 분리하여 정량화한 결과, 훈련데이터의 기간이 증가할수록 모델의 인식론적 불확실성이 감소하였으며, 훈련데이터 내의 이상치를 제거함으로써 내재적 불확실성이 감소하는 것을 확인할 수 있었다.

본 연구는 BNN을 활용하여 기계학습 모델 내에 존재하는 불확실성을 분리하여 정량화하는 방법을 소개한다. 모델의 예측 성능이 우수하더라도 훈련데이터의 양적, 질적 품질에 따라 불확실성의 크기가 달라질 수 있으며, 따라서 기계학습 모델의 성능을 검증하는 과정에서 예측 오

차뿐만 아니라 내재된 불확실성을 정량적으로 확인하는 것이 중요함을 본 논문을 통해 전달하고자 한다.

주요어 : 건물에너지 시뮬레이션, 기계학습, 베이지안 신경망, 인식론적 불확실성, 내재적 불확실성, 서포트 벡터 데이터 디스크립션
학 번 : 2018-28645

목 차

표 목 차

그 림 목 차

제 1 장 서론

1.1 연구 배경 및 목적

최근 건물에너지 관리시스템(Building Energy Management System, BEMS)의 도입의 확산으로 인해, 건물 전반에 대한 실내 환경, 에너지 소비 현황 및 설비 시스템의 운영 등에 대한 다양한 정보들을 수집할 수 있게 되었다. 수집된 데이터는 건물시스템의 동적 특성을 모사하는 시뮬레이션 모델을 제작하기 위해 활용될 수 있으며, 제작된 시뮬레이션 모델은 건물시스템에 대한 과학적이고 정량적인 분석에 활용될 수 있다.

건물에너지 시뮬레이션 모델은 동적 시뮬레이션 모델과 기계학습 모델이 주로 사용된다. 동적 시뮬레이션 모델은 열역학 제1 법칙에 기반하여 시간에 따라 변화하는 건물시스템의 동적 거동을 모사하며, 대표적으로 사용되는 도구로서 EnergyPlus, TRNSYS 등이 있다. 하지만, 지나치게 많은 입력변수를 요구하며, 일부 미지 변수에 대한 가정을 요구하기도 한다. 반면, 기계학습 모델은 물리 법칙에 기반을 두지 않고, 입력과 출력데이터 사이의 관계를 통계적으로 학습함으로써 건물의 동적 거동을 모사하는 모델이다. 기계학습 모델은 동적 시뮬레이션 모델에 비해 물리 시스템에 대한 전문적인 지식을 요구하지 않으며, 미지 변수에 대한 추정이 필요하지 않다. 이러한 장점으로, 기계학습은 건물에너지 시뮬레이션 분야에서 수요예측[1], 최적제어[2], 고장진단[3] 등 다양한 주제로 연구되

---

1) Afram, A., Janabi-Sharifi, F. (2014). Review of modeling methods for HVAC systems. Applied Thernal Engineering. 67, pp.507-519.

2) Afram, A., Janabi-Sharifi, F., Fung, A. S., Raahemifar, K. (2017). Artificial neural network based model predictive control and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system. Energy and Buildings. 141, pp.96-113.

3) Every, P. M. V., Rodriguez, M., Jones, C. B., Mammoli, A. A., Martínez-Ramón, M. (2017). Adavanced detection of HVAC faults using unsupervised SVM novelty detection and Gaussian process models. Energy and

고 있다. 하지만 기계학습 모델은 black-box 모델이기 때문에 입력과 출력 사이의 인과관계를 설명하는 데 제약이 있으며, 모델에 내재한 불확실성(uncertainty)이 존재한다.[4] 기계학습 모델의 불확실성은 모델의 예측 결과에 대한 신뢰도를 의미한다. 불확실성이 높은 모델은 모델 파라미터의 학습 및 선택 과정에 따라 매번 다른 결과를 출력할 수 있어, 모델의 실용적 사용에 제약이 될 수 있다. 따라서, 기계학습 모델을 안정적으로(robustly) 사용하기 위해서는 모델의 예측 성능뿐만 아니라 예측에 대한 불확실성을 정량적으로 분석하는 것이 중요하다.

인공신경망 (Artificial Neural Network, ANN), 순환 신경망(Recurrent Neural Network, RNN), 합성곱 신경망(Convolutional Neural Network, CNN) 등 일반적으로 사용되는 대부분의 딥러닝(deep learning) 알고리즘들은 결정적(deterministic) 결과를 출력하는 알고리즘으로, 모델의 불확실성에 대한 정보를 얻기 어렵다. 가우시안 프로세스 회귀(Gaussian Process Regression, GPR)는 모델의 신뢰도를 평가할 수 있는 대표적인 기계학습 알고리즘으로(그림 1-1), 입력에 대한 출력을 평균과 분산에 대한 함수의 형태로 학습하여 모델 예측에 대한 불확실성을 표현할 수 있다.[5] 하지만 GPR은 데이터와 파라미터의 수에 따라 대규모 행렬 연산이 포함될 수 있으며, 이는 연산 비용에 대한 GPR의 한계를 보여준다.

Buildings. 149, pp.216-224.

4) Kim, H., Jung, D. C., Choi, B. W. (2019). Exploiting the Vulnerability of Deep Learning-Based Artificial Intelligence Models in Medical Imaging: Adversarial Attacks. Journal of the Korean Society of Radiology, 80 (2):259

5) Williams, C. K., & Rasmussen, C. E. (2006). Gaussian processes for machine learning (Vol. 2). Cambridge, MA: MIT press.

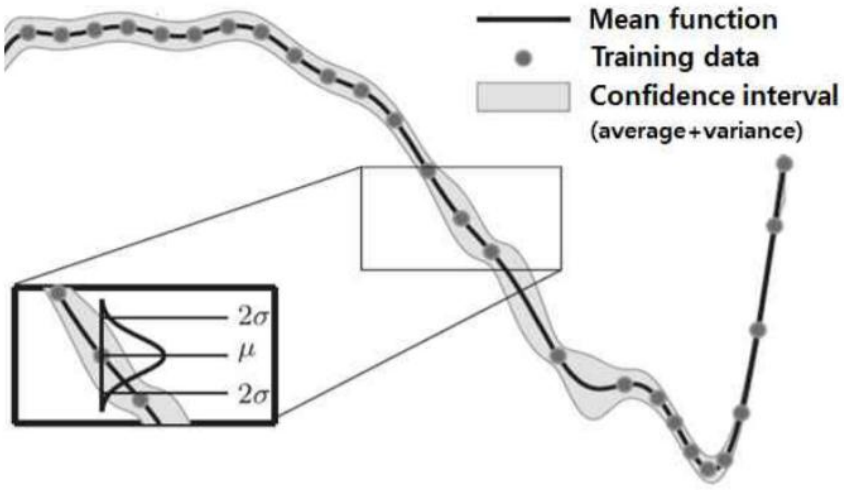


그림 1-1. 가우시안 프로세스 회귀(Gaussian Process Regression)[6)]

베이지안 신경망(Bayesian Neural Network, BNN)은 모델의 가중치(weight)에 대한 확률분포를 가정함으로써 확률적 해석을 가능하게 한 딥러닝 알고리즘이다(그림 1-2). 베이지안 신경망은 가중치 파라미터가 무수히 많아짐에 따라 신경망의 결과가 가우시안 프로세스로 수렴하게 되며[7)], 이는 GPR의 한계를 극복함과 동시에 모델 예측의 불확실성 또한 표현할 수 있음을 의미한다. 이러한 장점으로, 베이지안 신경망은 최근 전산과학 분야를 포함한 다양한 분야에서 주목받는 알고리즘이다.[8)]

6) 라선중, 신한솔, 서원준, 추한경, 박철수 (2017), 기존 건물 HVAC 시스템에 대한 다섯 가지 기계학습 모델 개발, 대한건축학회 논문집 제33권 제10호, pp.69-77

7) Neal, R. M. (1995). Bayesian learning for neural networks. PhD thesis, University of Toronto.

8) Kwon, Y., Won, J. H., Kim, B. J., & Paik, M. C. (2020). Uncertainty quantification using bayesian neural networks in classification: Application to biomedical image segmentation. Computational Statistics & Data Analysis, 142, 106816.

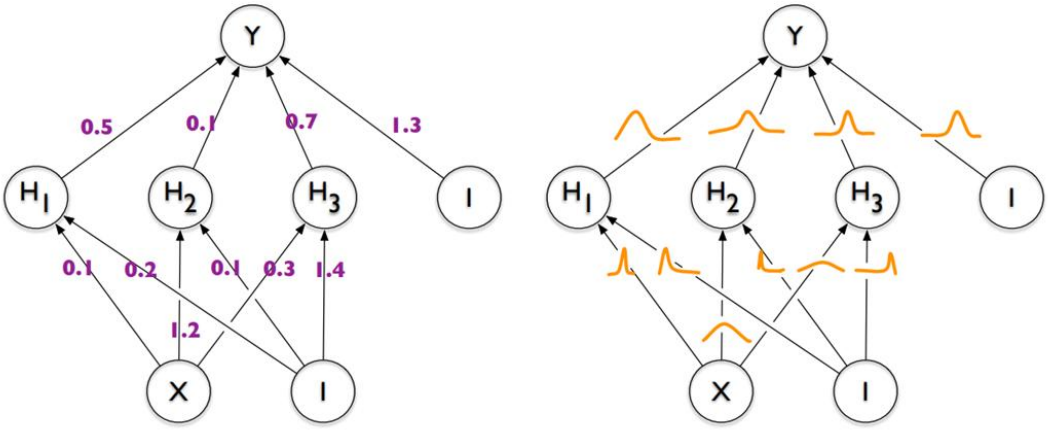


그림 1-2. 가중치가 고정값인 인공신경망(왼쪽)과 가중치에 대한 확률분포를 갖는 베이지안 신경망(오른쪽)[9)]

불확실성은 발생 원인에 따라 인식론적(epistemic) 불확실성과 내재적(aleatoric) 불확실성으로 구분된다. 먼저, 인식론적 불확실성은 'epistemic'이 의미하는 바처럼, 데이터 또는 지식의 부재(lack of data or knowledge)로 인해 발생하는 불확실성으로, 신경망 모델에서의 인식론적 불확실성은 신경망의 가중치에 의한 불확실성을 의미한다.[10)] 인식론적 불확실성은 1) 주어진 훈련데이터를 표현하는 모델 파라미터를 결정하기 어려운 경우, 또는 2) 훈련데이터의 범위를 벗어난 검증데이터를 사용한 경우에 주로 발생한다 (예: 여름철 데이터로 훈련된 모델을 겨울철에 사용하는 경우). 인식론적 불확실성은 모델 구조를 변경하거나, 추가적인 훈련데이터 수집을 통해 줄일 수 있다.[11)] 반면, 내재적 불확실성은 자연 현상의 본질적인(intrinsic) 무작위성(randomness)에 의해 발생한다. 즉, 기계학습

9) Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight uncertainty in neural networks. arXiv preprint arXiv:1505.05424.

10) Kendall, A. G. (2019). Geometry and uncertainty in deep learning for computer vision. PhD Thesis. University of Cambridge.

11) Der Kiureghian, A., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter?. Structural safety, 31(2), 105-112.; Nikolaidou, E., Wright, J., & Hopfe, C. J. (2015, December). Early and detailed design stage modelling using Passivhaus design; what is the difference in predicted building performance. 14th Conference of IBPSA, Vol. 9, pp. 2166-2173.

모델에서의 내재적 불확실성은 수집된 데이터 자체에 내재한(inherent) 노이즈(noise)나 이상치(outlier)에 의해 발생하는 불확실성이다. 이러한 노이즈 및 이상치는 센서의 센싱 오차 또는 시스템 자체의 오작동 등에 의해 발생할 수 있다. 센서를 교체하거나 시스템을 진단하는 등 데이터 측정 환경을 개선함으로써 내재적 불확실성을 줄일 수 있지만, 시간과 비용의 제약을 받을 수 있다. 따라서, 데이터의 전처리(pre-processing)를 통해 노이즈나 이상치를 제거함으로써 내재적 불확실성을 줄이는 방법이 사용된다.

본 논문에서는 베이지안 신경망을 이용하여 모델에 내재되어있는 불확실성을 정량화하는 방법을 소개하고, 훈련데이터의 양적, 질적 상태에 따라 변화하는 인식론적 및 내재적 불확실성의 크기를 비교하고자 한다. 이를 통해 기계학습 모델을 제작하고 적용하는 과정에서 불확실성의 정량적 평가에 대한 필요성을 제시하고, 추후 연구 방향에 대해 논의하고자 한다.

1.2 연구 범위 및 방법

본 논문에서는 베이지안 신경망 모델과 불확실성 정량화 방법을 통해, BEMS 데이터를 사용하여 기계학습 모델을 제작하는 과정에서 발생할 수 있는 불확실성에 대해 분석하였다. 베이지안 신경망은 연산량의 한계로 인하여 실용적인 구현이 어려우므로, Yarin Gal(2016)이 제안한 몬테카를로 드랍아웃(Monte Carlo Dropout, MC dropout) 방법을 사용하여 근사하였다. Y. Gal은 논문에서 드랍아웃이 적용된 인공신경망의 몬테카를로 근사를 통해 베이지안 신경망의 사후분포 추론과정을 모사할 수 있음을 주장하였으며, 이를 통해 신경망 모델의 불확실성을 정량화하는 방법을 제안하였다. 이후, Kendall&Gal(2017)[12]의 논문을 통해 베이지안 신경망의 불확실성을 인식론적 불확실성과 내재적 불확실성으로 분리하여

12) Kendall, A. G., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision?. In Advances in neural information processing systems (pp. 5574-5584).

정량화하는 방법을 제안하였다. 기계학습 모델의 불확실성은 훈련데이터의 양적, 질적 상태에 영향을 받기 때문에, 수집된 BEMS 데이터를 수집 기간, 이상치 검출 여부에 따라 4개의 훈련데이터 세트로 구분하였다. 이때, BEMS 데이터의 이상치를 식별하고 제거하기 위해, Tax&Duin(2004)[13]이 제안한 다변수 데이터의 이상치 검출 알고리즘인 서포트 벡터 데이터 디스크립션(Support Vector Data Description, SVDD)을 사용하였다. 4개의 훈련데이터 세트를 통해 4개의 베이지안 신경망 모델을 제작하였으며, 훈련데이터 이외의 모든 조건은 동일하게 구성하였다. 이후, 제작된 베이지안 신경망 모델은 검증 기간에 대한 예측 오차, 예측 재현성(reproducibility), 인식론적 및 내재적 불확실성의 측면에서 비교하였다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구에 사용된 이론들에 대한 배경 및 기본적인 수식 전개 과정에 대해 다룬다. 먼저, Y. Gal의 논문이 발표되기까지 베이지안 신경망을 구현하기 위한 기존 연구들을 소개하고, 해당 논문들의 이론적 기반이 되는 변분 추론(Variational Inference)에 관해 설명한다. 이후, Gal의 논문에 기반하여 MC dropout이 적용된 인공신경망과 베이지안 신경망이 동치를 이루는 과정을 설명하고, 베이지안 신경망의 출력결과를 통해 모델의 인식론적 불확실성과 내재적 불확실성을 분리하여 정량화하기 위한 Kendall의 방법론을 소개한다. 마지막으로, BEMS 데이터의 이상치를 검출 및 제거하는 과정에서 발생할 수 있는 이슈들과 이상치 검출 알고리즘인 SVDD에 대해 다룬다. 3장에서는 분석 대상인 업무시설의 야간 축냉용 압축식 냉동기에 대한 정보를 기술하고, 수집된 데이터의 이상치 제거 전후의 데이터 현황을 비교 분석하며, 베이지안 신경망의 훈련데이터와 파라미터 선정 과정을 서술한다. 4장에서는 제작된 베이지안 신경망 모델의 성능을 비교 분석하고, 5장의 결론을 통해 마무리한다.

13) Tax, D. M. J. and Duin, R. P. W. (2004). Support vector data description. Machine Learning, 54(1), (pp. 45–66).

제 2 장 베이지안 신경망과 모델 불확실성

본 장에서는 베이지안 신경망에 대한 개요와 관련 연구를 소개한다. 또한, MC dropout이 적용된 일반 인공신경망을 통해 베이지안 신경망의 결과를 모사하고, 이를 통해 모델의 인식론적 불확실성과 내재적 불확실성을 분리하여 정량화하는 방법을 소개한다.

2.1 베이지안 신경망 개요

베이지안 신경망은 모델의 가중치에 대해 사전분포(prior distribution)를 가정하고, 베이지안 추론과정을 거쳐 가중치에 대한 사후분포(posterior distribution)를 계산함으로써 신경망 모델의 확률적 해석을 가능하게 한 딥러닝 알고리즘이다. 입력데이터 $\mathrm{X}=\{\mathrm{x}_1,...,\mathrm{x}_\mathrm{N}\}$, 출력데이터 $\mathrm{Y}=\{\mathrm{y}_1,...,\mathrm{y}_\mathrm{N}\}$가 주어졌을 때, 베이지안 신경망의 가중치($\omega$)에 대한 사후분포 $p(\omega|\mathrm{X},\mathrm{Y})$는 사전분포 $p(\omega)$와 우도(likelihood) $p(\mathrm{Y}|\mathrm{X},\omega)$에 의해 계산된다(식 2.1). 신경망의 가중치가 확률분포로 표현되므로, 신경망을 통해 출력된 결과 또한 확률분포로 표현된다. 새로운 데이터 x^*가 입력으로 주어졌을 때, 베이지안 신경망의 출력 y^*는 사후예측분포(posterior predictive distribution)로 표현되며(식 2.2), 사후예측분포의 평균과 분산을 예측평균(predictive mean), 예측분산(predictive variance)이라 한다.

$$p(\omega|\mathrm{X},\mathrm{Y})=\frac{p(\mathrm{Y}|\mathrm{X},\omega)p(\omega)}{p(\mathrm{Y}|\mathrm{X})} \qquad \text{(식 2.1)}$$

$$p(\mathrm{y}^*|x^*,X,Y)=\int p(\mathrm{y}^*|x^*,\omega)p(\omega|\mathrm{X},\mathrm{Y})\,d\omega \qquad \text{(식 2.2)}$$

모델 증거(evidence)로 표현되는 식 2.1의 $p(Y|X)$는 우도의 주변화(周邊化) 방법(marginalization)로 계산된다(식 2.3). 하지만, 딥러닝 모델과 같이 수많은 불확실한 파라미터가 존재할 경우, 모든 파라미터 공간에 대해 적분하는 것은 어렵다.

$$p(\mathrm{Y}|\mathrm{X}) = \int p(\mathrm{Y}|\mathrm{X},\omega)p(\omega)\,d\omega \qquad \text{(식 2.3)}$$

베이지안 신경망의 이론적 배경은 비교적 간단함에도 불구하고, 과도한 연산량을 요구하여 실용적인 구현은 어렵다고 인식되어 왔다.[14] 하지만, 최근 연산기술의 발달과 함께 베이지안 신경망을 실용적으로 구현하는 다양한 방법론이 제시되고 있다. 제안된 방법론은 대표적으로 샘플링 기반 방법과 변분추론 등이 있다. 이어지는 절에서는 변분추론이 적용된 베이지안 신경망 연구들의 소개와 함께 변분추론에 대해 설명하고자 한다.

2.2 베이지안 신경망 및 변분추론

2.2.1 관련 연구

Hinton&Van Camp(1993)[15]는 베이지안 신경망의 가중치에 대한 사후분포 추정과정을 정보이론의 기본 개념 중 하나인 최소 설명 길이(Minimum description length, MDL)를 기반으로 설명하였다. MDL의 개념 아래에서, 가장 좋은 모델은 모델과 데이터 사이의 차이(misfits)와 모델의 가중치(weights)를 설명하는 길이(description length)를 최소화하는 모델

14) Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In international conference on machine learning (pp. 1050-1059)

15) Hinton, G. E., & Van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. In Proceedings of the sixth annual conference on Computational learning theory (pp. 5-13).

로 표현된다. 해당 논문에서는 입력정보와 정답(correct output)에 대해 알고 있는 발신자(sender)와 입력정보만을 알 수 있는 수신자(receiver)를 예시로 신경망에서의 MDL을 설명한다. 발신자와 수신자가 모델 가중치에 대한 사전분포 P를 공유하고 있을 때, 발신자가 알고 있는 사후분포 Q에 대한 정보를 수신자에게 전달하기 위해 요구되는 설명 길이(expected description length) 혹은 통신 비용(communication cost)은 식 2.4로 표현하며, 이를 최소화하는 것을 목적으로 한다. Hinton&Van Camp의 방법론은 후술할 변분추론의 목적과 과정을 매우 닮아, 처음으로 변분추론(variational inference)의 개념이 도입된 베이지안 신경망으로 평가하기도 한다.

$$G(P, G) = \int Q(\omega) \log \frac{Q(\omega)}{P(\omega)} d\omega \qquad \text{(식 2.4)}$$

Hinton&Van Camp(1993)의 논문에서는 MDL의 개념 아래에서 단층 인공신경망의 가중치에 대한 사후분포를 추정하였다. 이 과정에서 계산의 단순화를 위해 신경망의 가중치가 서로 독립적임을 가정하였으며, 가우시안 추정분포의 분산을 대각행렬(diagonal matrix)로 표현하였다. 이에 대한 보완으로, Barber&Bishop(1998)[16]은 가중치 사이의 상관관계를 설명하기 위해, 사후분포를 공분산(covariance)을 가지는 결합분포(joint distribution) 형태로 추정하였다. 하지만, 상기 두 방법 모두 한 개의 은닉층을 가진 순방향 신경망(feed forward neural network)을 대상으로 적용되었으며, 훈련데이터의 양이 많거나, 딥러닝 알고리즘(CNN, RNN 등)과 같이 복잡한 구조를 가진 신경망에는 적용이 어렵다는 한계가 있다.[17]

Gal(2016)은 모든 층(layer)에 드랍아웃이 적용된 일반적인 인공신경망의 학습 과정과 변분추론을 통해 베이지안 신경망의 사후분포를 추정하

16) Barber, D., & Bishop, C. M. (1998). Ensemble learning in Bayesian neural networks. Nato ASI Series F Computer and Systems Sciences, 168, (pp. 215-238).
17) Gal, Y. (2016). Uncertainty in deep learning. PhD Thesis. University of Cambridge.

는 과정이 동일함을 증명하였다. 또한, 드랍아웃 신경망의 몬테카를로 근사(MC dropout)를 통해 사후예측분포를 근사하고, 불확실성을 정량화하는 방법을 제안하였다. Gal의 방법론은 은닉층이 두 개 이상인 인공신경망에도 적용할 수 있으며, 순환 신경망, 합성곱 신경망 등 딥러닝 알고리즘에도 쉽게 적용할 수 있다는 장점이 있다.

2.2.2 변분추론

본 절에서는 Gal의 논문들[18]을 기반으로 변분추론에 대한 소개와 함께 베이지안 신경망의 추론이 드랍아웃 인공신경망의 학습과 동치를 이루는 과정에 대해 기술한다. 변분추론은 분석적(analytical) 계산이 어려운 사후분포를 추정하기 위해 비교적 계산이 단순한 추정분포(예: 가우시안(Gaussian), 감마(gamma), 베르누이(Bernoulli) 분포 등)를 통해 근사하는 방법이다. 추정분포는 $q_\theta(\omega)$로 표현하며, 파라미터 θ에 의해 조절된다. 변분추론의 목적은 사후분포와 가장 유사한 추정분포를 찾는 것이므로, 서로 다른 두 확률분포 사이의 유사도를 계산하는 쿨백-라이블러 발산(Kullback-Leibler Divergence, KLD)을 이용하여 사후분포와 추정분포 사이의 유사도를 계산한다(식 2.5).

$$KL(q_\theta(\omega) \parallel p(\omega|\mathrm{X}, \mathrm{Y})) = \int q_\theta(\omega) \log \frac{q_\theta(\omega)}{p(\omega|\mathrm{X}, \mathrm{Y})} d\omega \quad \text{(식 2.5)}$$

사후분포와 추정분포 간의 KLD를 최소화하는 문제는 식 2.6으로 표현되는 ELBO(Evidence Lower Bound)를 최대화(혹은 negative ELBO를 최소화)하는 문제와 같다. 결국, 사후분포의 계산을 위한 적분 과정이 변분

18) Gal, Y. (2016). Uncertainty in deep learning. PhD Thesis. University of Cambridge.; Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In international conference on machine learning (pp. 1050-1059).

추론의 적용을 통해 최적화 문제로 전환되며, 이를 통해 신경망을 학습하는 과정을 베이지안 역전파(Bayes by Backprop, BBB)라 한다. 결과적으로, 베이지안 신경망의 가중치 학습을 위한 목적함수는 식 2.7로 표현되며, N은 훈련데이터 수, M은 미니 배치(mini-batch)의 크기, S는 미니 배치의 인덱스를 의미한다.

$$ELBO = \int q_\theta(\omega) \log p(Y|X,\omega) d\omega - KL(q_\theta(\omega) \parallel p(\omega)) \quad \text{(식 2.6)}$$

$$\widehat{L_{VI}}(\theta) := -\frac{N}{M} \sum_{i \in S} \int q_\theta(\omega) \log p(\mathrm{y}_i|\mathrm{f}^w(\mathrm{x}_i)) d\omega \quad \text{(식 2.7)}$$

$$+ KL(q_\theta(\omega) \parallel p(\omega))$$

2.2.3 재매개변수화(re-parametrization trick)

베이지안 역전파를 수행하기 위해서는 추정분포의 파라미터인 θ에 대한 ELBO의 기울기(gradient)를 구하고, 기울기에 따라 기존의 θ를 업데이트한다. 하지만, ELBO의 계산 과정 또한 적분을 포함하고 있어, 여전히 신경망이 두 층 이상일 경우에는 계산이 어렵다. 따라서, 몬테카를로 샘플링을 통해 ELBO를 근사하는 기법들이 사용된다. 식 2.8은 ELBO의 기울기를 계산하는 과정에서 적분항만 나타낸 것으로, 로그 우도에 대한 기댓값(expectation)의 기울기를 의미한다. 여기서 $q_\theta(\omega)$는 θ에 대해 종속적인 분포로서, 단순히 $q_\theta(\omega)$ 분포에서 샘플링하여 근사할 수 없다. 따라서, $q_\theta(\omega)$를 θ에 독립적인 분포로 변환한다면 상기 문제를 해결할 수 있다. 만일, 가중치 ω를 θ와 ϵ에 대해 재매개변수화(re-parameterization) 함으로써(식 2.9), 식 2.8은 식 2.10-a로 변환될 수 있고, 이는 $p(\epsilon)$에 대한 기댓값 형태로 표현 가능하며(식 2.10-b), 따라서 몬테카를로 샘플링을 통

해 근사할 수 있다(식 2.10-c).

$$\frac{\delta}{\delta\theta}\int q_\theta(w)\log p(\mathrm{y}|\mathrm{f}^w(\mathrm{x}))dw \qquad \text{(식 2.8)}$$

$$\omega = g(\theta,\epsilon) \qquad \text{(식 2.9)}$$

$$\frac{\delta}{\delta\theta}\int p(\epsilon)\log p(\mathrm{y}|\mathrm{f}^{g(\theta,\epsilon)}(\mathrm{x}))d\epsilon \qquad \text{(식 2.10-a)}$$

$$= \mathrm{E}_{p(\epsilon)}[\frac{\delta}{\delta\theta}\log p(\mathrm{y}|\mathrm{f}^{g(\theta,\epsilon)}(\mathrm{x}))] \qquad \text{(식 2.10-b)}$$

$$\approx \frac{1}{\mathrm{T}}\sum_{\mathrm{t}=1}^{\mathrm{T}}\frac{\delta}{\delta\theta}\log p(\mathrm{y}|\mathrm{f}^{g(\theta,\epsilon_t)}(\mathrm{x})) \qquad \text{(식 2.10-c)}$$

결과적으로, 변분추론과 재매개변수화를 이용한 베이지안 신경망의 추론 과정은 식 2.11로 표현되며, 여기서 $\hat{\epsilon_i}$는 $p(\epsilon)$에서 무작위로 추출된 변수를 의미한다.

$$\frac{\delta}{\delta\theta}\hat{L}_{VI}(\theta) = -\frac{N}{M}\sum_{i\in S}\frac{\delta}{\delta\theta}\log p(\mathrm{y}_i|\mathrm{f}^{g(\theta,\hat{\epsilon_i})}(\mathrm{x}_i)) \qquad \text{(식 2.11)}$$

$$+\frac{\delta}{\delta\theta}KL(q_\theta(\omega)\parallel p(\omega))$$

2.3 드랍아웃 인공신경망

드랍아웃은 신경망의 과적합(overfitting)을 방지하기 위해 제안된 인공신경망의 정규화(regularization) 방법으로, 일부 노드를 탈락시키면서 학습하는 방법이다.[19] 신경망의 모든 층에 드랍아웃을 적용하는 것은 베이지안 신경망의 학습 과정과 동일함이 다음과 같이 증명되었다 (Gal, 2016). 먼저, 신경망 각 층에 드랍아웃을 적용하기 위해 $\hat{\epsilon}_1$과 $\hat{\epsilon}_2$를 샘플링한다. $\hat{\epsilon}_1$, $\hat{\epsilon}_2$는 이진 벡터로, 입력변수의 차원인 Q, 은닉층의 차원인 K와 동일한 차원을 가지며, $\hat{\epsilon}_i$의 요소들은 각각 p_1과 p_2의 확률로 0의 값을 갖는 베르누이 분포(Bernoulli distribution)에서 추출된 값이다. 입력변수 x가 주어지면, 첫 번째 드랍아웃으로 인해 은닉층으로 전달되는 실제 입력은 x와 $\hat{\epsilon}_1$의 대각행렬의 곱인 $\hat{\mathrm{x}}=\mathrm{x}(diag(\hat{\epsilon}_1))$이다. 입력층과 은닉층 사이의 가중치 행렬을 M_1, 은닉층과 출력층 사이의 가중치 행렬을 M_2, 은닉층의 편향(bias)을 b라 할 때, 은닉층에 전달된 입력값 $\hat{\mathrm{x}}$이 활성함수(activation function) σ를 거쳐 출력된 값은 $\mathrm{h}=\sigma(\hat{\mathrm{x}}\mathrm{M}_1+\mathrm{b})$이다. 두 번째 드랍아웃을 거쳐 $\hat{\mathrm{h}}=\mathrm{h}(diag(\hat{\epsilon}_2))$이 출력층으로 전달되며, 최종 신경망의 출력값은 $\hat{\mathrm{y}}=\hat{\mathrm{h}}\mathrm{M}_2$이 된다(식 2.12). 신경망의 가중치 벡터 $\hat{\mathrm{W}}_1$, $\hat{\mathrm{W}}_2$을 $\hat{\mathrm{W}}_1 := diag(\hat{\epsilon}_1)\mathrm{M}_1$, $\hat{\mathrm{W}}_2 := diag(\hat{\epsilon}_2)\mathrm{M}_2$로 정의하면, 신경망의 최종 출력값은 식 2.13으로 표현된다.

$$\hat{\mathrm{y}}=\sigma(\mathrm{x}(diag(\hat{\epsilon}_1)\mathrm{M}_1)+\mathrm{b})(diag(\hat{\epsilon}_2)\mathrm{M}_2) \qquad \text{(식 2.12)}$$

19) Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.; Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

$$\hat{\mathrm{y}} = \sigma(\mathrm{x}\widehat{\mathrm{W}}_1 + \mathrm{b})\widehat{\mathrm{W}}_2 =: \mathrm{f}^{\widehat{\mathrm{W}}_1, \widehat{\mathrm{W}}_2, \mathrm{b}}(\mathrm{x}) \qquad \text{(식 2.13)}$$

드랍아웃의 적용으로 인해 무작위로 추출된 신경망의 가중치 집합을 $\hat{\omega}_i$으로 정의하면(식 2.14-a, 2.14-b), $\hat{\omega}_i$는 θ와 $\hat{\epsilon}_i$에 대한 함수로 나타낼 수 있다(식 2.14-c). 여기서, θ는 드랍아웃이 적용되지 않은 신경망의 가중치 집합 $\theta = \{\mathrm{M}_1, \mathrm{M}_2, \mathrm{b}\}$를 의미한다.

$$\hat{\omega}_i = \left\{\widehat{\mathrm{W}}_1^i, \widehat{\mathrm{W}}_2^i, \mathrm{b}\right\} \qquad \text{(식 2.14-a)}$$

$$= \left\{diag(\hat{\epsilon}_1^i)\mathrm{M}_1, diag(\hat{\epsilon}_2^i)\mathrm{M}_2, \mathrm{b}\right\} \qquad \text{(식 2.14-b)}$$

$$= g(\theta, \hat{\epsilon}_i) \qquad \text{(식 2.14-c)}$$

드랍아웃이 적용된 신경망의 목적함수는 식 2.15로 표현된다. 여기서, M은 미니 배치의 크기, τ^{-1}는 모델의 관측 오차(observation noise), λ는 과적합 방지를 위한 가중치 감소 상수(weight-decay)이다. 결과적으로, 베이지안 신경망의 재매개변수화를 이용한 변분추론 과정(식 2.11)은 드랍아웃을 적용한 인공신경망의 최적화 과정(식 2.15)과 동일함을 알 수 있다.

$$\frac{\delta}{\delta\theta}\hat{L}_{d.o.}(\theta) = -\frac{1}{\mathrm{M}\tau}\sum_{i\in S}\frac{\delta}{\delta\theta}\log p(\mathrm{y}_i|\mathrm{f}^{g(\theta,\hat{\epsilon}_i)}(\mathrm{x}_i)) \qquad \text{(식 2.15)}$$

$$+\frac{\delta}{\delta\theta}\lambda(\|\mathrm{M}_1\|^2 + \|\mathrm{M}_2\|^2 + \|\mathrm{b}\|^2)$$

2.4 모델 불확실성

변분추론 과정을 통해 최적화된 추정분포를 적용함으로써, 사후예측 분포(식 2.3)는 식 2.16과 같이 변형된다. 여기서, $\omega = \{\mathrm{W}_i\}_{i=1}^{L}$ 은닉층 L 에서 추출된 무작위 가중치, $\mathrm{f}^{\omega}(\mathrm{x}^*)$는 신경망의 출력값, $q_\theta^*(\omega)$는 변분추론 과정(식 2.7)을 통해 최적화된 추정분포를 의미한다.

$$q_\theta^*(\mathrm{y}^*|\mathrm{x}^*) = \int p(\mathrm{y}^*|\mathrm{f}^{\omega}(\mathrm{x}^*)) q_\theta^*(\omega)\, d\omega \qquad \text{(식 2.16)}$$

베이지안 신경망 모델의 불확실성은 모델 출력 y^*의 예측분산을 의미하며, 식 2.17로 정의된다. y^* 및 $(\mathrm{y}^*)^{\mathrm{T}}(\mathrm{y}^*)$의 기댓값(식 2.18-a, 2.19-a)은 신경망의 출력값 $\mathrm{f}^{\hat{\omega}_t}(\mathrm{x}^*)$의 몬테카를로 근사를 통해 구할 수 있으며(식 2.18, 2.19), 결과적으로 y^*의 예측분산은 식 2.20으로 계산할 수 있다. 여기서, $\hat{w}_t$는 $q_\theta^*(\omega)$에서 무작위로 추출된 가중치, τ^{-1}는 모델의 예측 정확성을 의미한다.

$$\mathrm{Var}[\mathrm{y}^*] = \mathrm{E}[(\mathrm{y}^*)^{\mathrm{T}}(\mathrm{y}^*)] - \mathrm{E}[\mathrm{y}^*]^{\mathrm{T}}\mathrm{E}[\mathrm{y}^*] \qquad \text{(식 2.17)}$$

$$\mathrm{E}_{q_\theta(y^*|x^*)}[\mathrm{y}^*] = \int \mathrm{y}^* q_\theta^*(y^*|x^*) d\mathrm{y}^* \qquad \text{(식 2.18-a)}$$

$$= \int\int y^* p(\mathrm{y}^*|\mathrm{f}^{\omega}(\mathrm{x}^*)) q_\theta^*(\omega) d\omega d\mathrm{y}^* \qquad \text{(식 2.18-b)}$$

$$= \int f^{\omega}(x^*) q_\theta^*(\omega) d\omega \qquad \text{(식 2.18-c)}$$

$$\approx \frac{1}{T}\sum_{t=1}^{T} \mathrm{f}^{\hat{\omega}_t}(\mathrm{x}^*) \qquad \text{(식 2.18-d)}$$

$$\mathrm{E}_{q_\theta(y^*|x^*)}[(\mathrm{y}^*)^\mathrm{T}(\mathrm{y}^*)] = \int (\mathrm{y}^*)^\mathrm{T}(\mathrm{y}^*) q_\theta^*(y^*|x^*) d\mathrm{y}^* \quad \text{(식 2.19-a)}$$

$$= \int \left(\int (y^*)^\mathrm{T}(y^*) p(\mathrm{y}^*|\mathrm{f}^\omega(\mathrm{x}^*)) d\mathrm{y}^* \right) q_\theta^*(\omega) d\omega \quad \text{(식 2.19-b)}$$

$$= \int \left(\tau^{-1}\mathrm{I} + \mathrm{f}^\omega(\mathrm{x}^*)^\mathrm{T}\mathrm{f}^\omega(\mathrm{x}^*) \right) q_\theta^*(\omega) d\omega \quad \text{(식 2.19-c)}$$

$$\approx \tau^{-1}\mathrm{I} + \frac{1}{T}\sum_{\mathrm{t}=1}^{T} \mathrm{f}^{\hat{\omega}_t}(\mathrm{x}^*)^\mathrm{T}\mathrm{f}^{\hat{\omega}_t}(\mathrm{x}^*) \quad \text{(식 2.19-d)}$$

$$\widetilde{\mathrm{Var}}[\mathrm{y}^*] = \tau^{-1}\mathrm{I} + \frac{1}{T}\sum_{\mathrm{t}=1}^{T} \mathrm{f}^{\hat{\omega}_t}(\mathrm{x}^*)^\mathrm{T}\mathrm{f}^{\hat{\omega}_t}(\mathrm{x}^*) - \frac{1}{T^2}\left(\sum_{\mathrm{t}=1}^{T} \mathrm{f}^{\hat{\omega}_t}(\mathrm{x}^*) \right)^\mathrm{T} \left(\sum_{\mathrm{t}=1}^{T} \mathrm{f}^{\hat{\omega}_t}(\mathrm{x}^*) \right) \quad \text{(식 2.20)}$$

회귀모델에서 일반적으로 사용되는 가우시안 우도는 모델 정확성을 의미하는 τ^{-1}로 표현된다(식 2.21). 여기서 τ^{-1}는 모델의 정확한 예측을 방해하는 관측 오차(observation noise)로 해석할 수 있다. 결국, τ^{-1}는 데이터 자체에 내재된 불확실성, 즉 '내재적 불확실성'을 의미한다.

$$p(\mathrm{y}|\mathrm{x},\omega) = N(\mathrm{y}; \mathrm{f}^\omega(\mathrm{x}), \tau^{-1}\mathrm{I}) \quad \text{(식 2.21)}$$

사용자가 우도를 정의할 때, τ^{-1}를 상수로 가정하면, 입력데이터가 변화하여도 불확실성이 변하지 않고 일정한 값으로 고정된 등분산적(homoscedastic) 내재적 불확실성을 갖는다. 반면 τ^{-1}를 상수로 정하지 않고 데이터를 통해 학습하면, 입력데이터의 변화에 따른 불확실성을 구할 수 있으며, 이는 이분산적(heteroscedastic) 내재적 불확실성으로 정의된다. Kendall&Gal(2017)[20]에서는 모델의 내재적 불확실성을 정량화하기 위해,

20) Kendall, A. G., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision?. In Advances in neural information processing

데이터를 통해 모델의 정확성을 학습하는 인공신경망 모델을 제작하였다. 해당 인공신경망은 은닉층이 분리되어 모델의 결과로 평균($\hat{y}$)과 분산($\hat{\sigma}^2$)을 출력하며(식 2.22), $\hat{y}$와 $\hat{\sigma}^2$가 함께 손실함수(loss function)에 포함되어 학습되기 때문에(식 2.23), 분산($\hat{\sigma}^2$)에 대한 별도의 레이블(label) 없이도 학습할 수 있다.

$$[\hat{y}, \hat{\sigma}^2] = f^{\omega}(x) \quad \text{(식 2.22)}$$

$$Loss = \sum_i \frac{1}{2\hat{\sigma}_i^2} \| y_i - \hat{y}_i \|^2 + \frac{1}{2}\log \hat{\sigma}_i^2 \quad \text{(식 2.23)}$$

새롭게 학습된 인공신경망의 출력과 함께, 식 2.20의 예측분산은 식 2.24로 계산 가능하며, 여기서 T는 샘플링 횟수, $\hat{y_t}$와 $\hat{\sigma_t}^2$는 t번째 샘플링된 신경망 모델의 출력을 의미한다. 즉, 최종적인 베이지안 신경망의 불확실성을 의미하며, 인식론적 불확실성(식 2.25-a, $U_{epistemic}$)과 내재적 불확실성(식 2.25-b, $U_{aleatoric}$)으로 분리할 수 있다. 결과적으로, 드랍아웃이라는 확률적 처리 과정을 통해 베이지안 신경망의 추론과정을 모사할 수 있으며, 신경망 결과의 몬테카를로 근사를 통해 신경망 모델의 인식론적 불확실성과 내재적 불확실성을 분리하여 정량화할 수 있다.

$$\widetilde{Var}[\hat{y}] = \frac{1}{T}\sum_{t=1}^{T} \hat{y}_t^2 - \left(\frac{1}{T}\sum_{t=1}^{T} \hat{y}_t\right)^2 + \frac{1}{T}\sum_{t=1}^{T} \hat{\sigma}_t^2 \quad \text{(식 2.24)}$$

$$U_{epistemic} : \frac{1}{T}\sum_{t=1}^{T} \hat{y}_t^2 - \left(\frac{1}{T}\sum_{t=1}^{T} \hat{y}_t\right)^2 \quad \text{(식 2.25-a)}$$

$$U_{aleatoric} : \frac{1}{T}\sum_{t=1}^{T} \hat{\sigma}_t^2 \quad \text{(식 2.25-b)}$$

---

systems (pp. 5574-5584).

2.5 이상치 검출 알고리즘

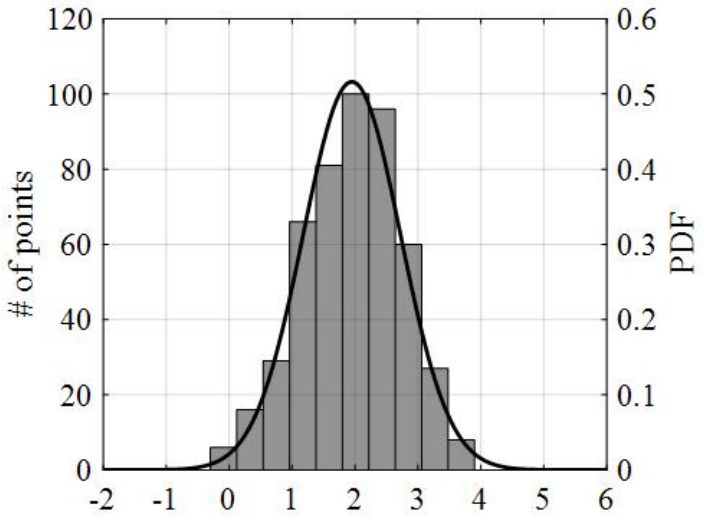
2.5.1 BEMS 데이터 이상치 검출의 쟁점

기계학습 모델은 데이터의 입력과 출력 사이의 관계를 통계적으로 학습하기 때문에, 훈련데이터의 양적, 질적 품질에 따라 모델의 성능이 좌우된다. 만일, 훈련데이터 내에 존재하는 비정상 데이터를 제거하지 않고 기계학습 모델을 제작할 경우, 정상적인 시스템의 거동을 모사하기 어려우며, 모델 예측 정확도가 저하될 수 있다. 따라서, 기계학습 모델을 제작하기 전 BEMS 데이터 내 비정상 데이터에 대한 식별 및 제거 과정을 수행하는 것이 중요하다.

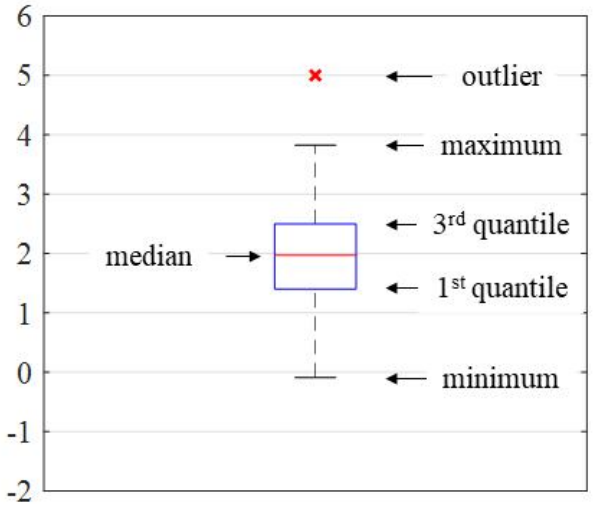
데이터의 이상치 처리는 비정상 상태에 대한 레이블(label)의 수집 가능 여부에 따라 지도 학습(supervised learning) 방식과 비지도 학습(unsupervised learning) 방식으로 구분된다. 만일, BEMS를 통해 수집된 데이터에 대해 시간대별로 시스템의 고장 여부나 이상 상태를 확인할 수 있으며 이상 데이터를 충분히 확보할 수 있다는 가정하에, 지도 학습 알고리즘을 사용하여 시스템의 고장 여부를 학습하는 모델을 제작할 수 있다. 하지만, 매시간 HVAC 시스템의 이상 상태를 진단하는 것은 매우 비용 소모적이므로, 일반적인 BEMS에서는 HVAC 시스템의 이상 상태에 대한 레이블을 측정하지 않는다. 따라서, BEMS 데이터의 이상치 검출을 위해서는 이상 상태에 대한 별도의 레이블을 요구하지 않는 비지도 학습 알고리즘이 주로 사용된다.

가장 일반적으로 사용되는 비지도 이상치 검출 방법은 데이터의 확률밀도를 이용한 방법으로(그림 2-1), 다음과 같은 과정을 통해 수행된다. 먼저, 확률 모델(정규분포, 히스토그램, 박스 플롯 등)을 이용하여, 데이터의 확률밀도를 추정한다. 다음으로, 분석자는 추정된 확률밀도에 대해 정상 데이터의 임계치(threshold)를 정의하고, 정의된 임계치를 넘어서는 데이터를 비정상 데이터로 분류하게 된다. 예를 들어, 데이터의 확률밀도를 정규분포로 추정한 경우에는 표준편차를, 박스 플롯의 경우

IQR(Interquartile Range)를 이용하여 확률 모델의 임계치를 표현한다.



(a) 히스토그램, 정규분포

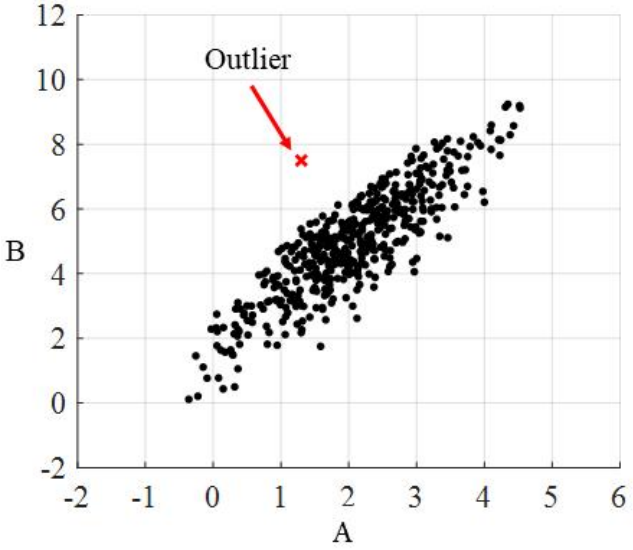


(b) 박스 플롯

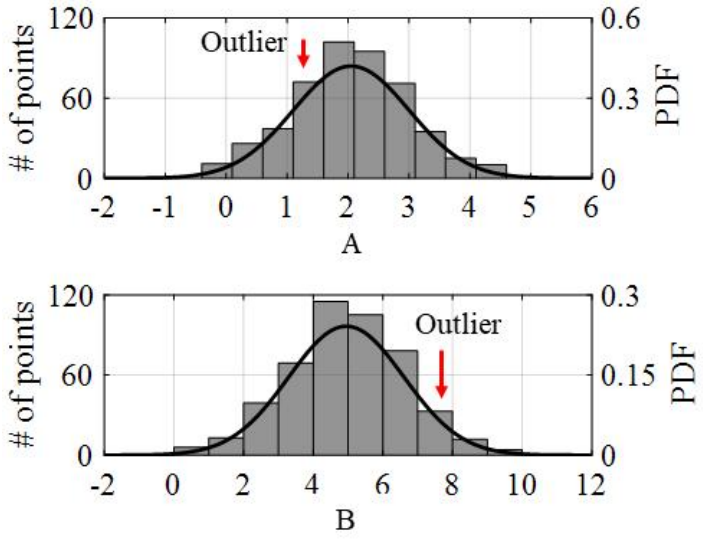
그림 2-1. 비지도 이상치 검출 방법(단변수)

하지만, 이와 같은 이상치 처리 방식에는 한계가 존재한다. 예를 들어, 아래의 산점도(scatter-plot)에서 정상 데이터(검은색 •)와 달리 눈에 띄게 벗어난 비정상 데이터(붉은색 ×)를 시각적으로 확인할 수 있다(그

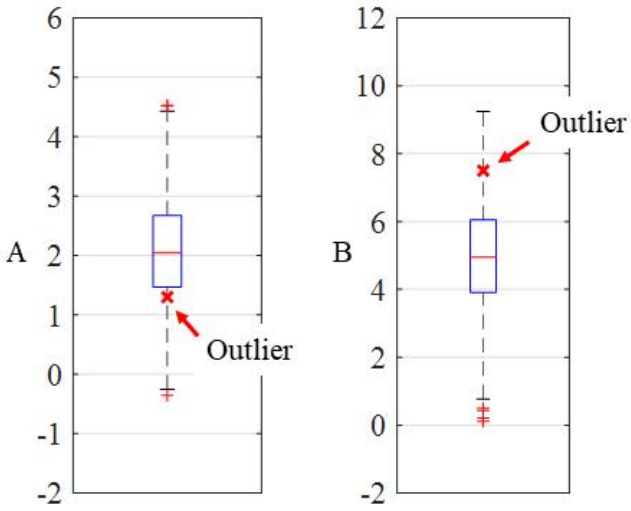
림 2-2(a)). 하지만, 상기 언급한 방법과 같이 각 변수 A와 B에 대한 확률 모델을 작성했을 때, 산점도에서는 확인할 수 있었던 비정상 데이터가 정상 데이터의 범위 내에 존재함을 알 수 있다(그림 2-2(b), (c)). 즉, 데이터의 변수가 2개 이상일 경우에는 각 변수에 대한 통계 처리만으로는 비정상 데이터를 정확히 식별하기 어려울 수 있으며, 변수 간의 상관관계를 고려한 다변수 이상치 판별 방법을 적용하는 것이 필요하다.



(a) 산점도



(b) 히스토그램, 정규분포



(c) 박스 플롯

그림 2-2. 비지도 이상치 검출 방법(다변수)

요약하면, BEMS 데이터의 이상치 처리 과정이 단순하지 않다는 것을 알 수 있다. 먼저, BEMS 데이터 자체에 이상 상태를 알 수 있는 레이블이 존재하지 않으므로, 비지도 학습 방법을 이용하여 이상 데이터를 식별해야 한다. 일반적으로 비지도 학습은 정답, 즉 이상 상태에 대한 레이블을 알 수 없으므로, 데이터를 통해 추정해야 하며 식별 결과에 대한 검증이 어렵다. 또한, 데이터의 변수가 2개 이상인 다변수 데이터의 이상치를 제거할 때 개별 변수에 대한 확률 모델링만으로는 정확한 식별이 어렵다. 따라서, 변수 간의 상관관계를 고려하는 방법을 사용해야 하지만, 데이터의 형태에 따라 처리 과정이 매우 어려워질 수 있다.

2.5.2 Support Vector Data Description

SVDD는 비지도 분류기(unsupervised classifier)의 하나이며, 정상 데이터를 감싸는 최소 체적의 초구(hypersphere)를 탐색하고, 이를 통해 정상 데이터의 경계를 학습하여, 학습된 경계를 통해 정상 데이터와 비정상

데이터를 구분한다. SVDD의 학습 과정은 다음과 같다. N개의 훈련데이터 $\mathrm{X}=\{\mathrm{x}_1,...,\mathrm{x}_N\}$가 주어졌을 때, 데이터의 경계를 나타내는 초구의 중심(a)과 반지름(R)은 제약조건이 있는 최소화 문제를 통해 구할 수 있다(그림 2-3, 식 2.26). 데이터의 오차를 표현하기 위해 구의 범위 밖에 존재하는 데이터에 대해서는 패널티(ξ, slack variable)를 부과하며, 파라미터(C)로 초구체의 체적과 패널티 사이의 균형(trade-off)을 조절한다.

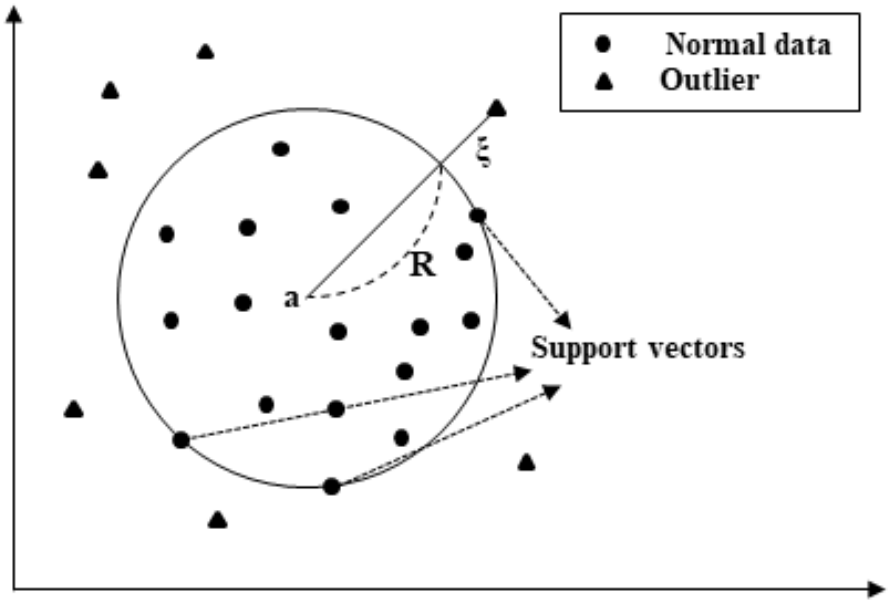


그림 2-3. Support Vector Data Description

$$MIN\ \mathrm{f}(\mathrm{R},\mathrm{a}) = \mathrm{R}^2 + \mathrm{C}\sum_i \xi_i \qquad \text{(식 2.26)}$$
$$s.t.\ \| x_i - \mathrm{a} \|^2 \leq \mathrm{R}^2 + \xi_i\ ,\ \xi_i \geq 0,\ \forall\, i$$

라그랑주 승수(Lagrange multiplier)인 α_i와 γ_i를 도입함으로써 식 2.26의 제약조건과 최소화 문제는 식 2.27과 같이 라그랑주 쌍대 문제(Lagrangian dual problem)로 구성할 수 있다.

$$L(\mathrm{R},\mathbf{a},\alpha_i,\gamma_i,\xi_i) = \mathrm{R}^2 + \mathrm{C}\sum_i \xi_i \qquad \text{(식 2.27)}$$
$$-\sum_i \alpha_i \left\{ \mathrm{R}^2 + \xi_i - (\| x_i \|^2 - 2\mathbf{a} \bullet x_i + \| \mathbf{a} \|^2) \right\} - \sum_i \gamma_i \xi_i$$
단, $(\alpha_i \geq 0\ ,\ \gamma_i \geq 0)$

식 2.27의 함수 L이 최소가 되는 지점은 각 변수 R, a, ξ_i에 대해 편미분 0인 지점에서 존재할 수 있으며, 이는 새로운 제약조건을 제공한다 (식 2.28-2.30).

$$\frac{\partial L}{\partial R} = 0 : \quad \sum_i \alpha_i = 1 \qquad \text{(식 2.28)}$$

$$\frac{\partial L}{\partial \mathbf{a}} = 0 : \quad \mathbf{a} = \frac{\sum_i \alpha_i x_i}{\sum_i \alpha_i} = \sum_i \alpha_i x_i \qquad \text{(식 2.29)}$$

$$\frac{\partial L}{\partial \xi_i} = 0 : \quad C - \alpha_i - \gamma_i = 0 \qquad \text{(식 2.30)}$$

세 가지 제약조건 아래에서 식 2.27의 최소화 함수 L은 식 2.31과 같이 표현된다.

$$L = \sum_{i,j} \alpha_i \alpha_j (x_i \bullet x_j) - \sum_i \alpha_i (x_i \bullet x_i) \qquad \text{(식 2.31)}$$

식 2.31을 최대화하는 α_i를 구함으로써 초구체의 경계를 구분하는 서포트 벡터(support vector)가 결정된다. Karush-Kuhn-Tucker 조건으로 모든 i에 대해 식 2.32을 만족해야 하며, $\alpha_i > 0$일 경우 $\| \mathrm{x}_i - \mathbf{a} \|^2 = \mathrm{R}^2 + \xi_i$, $\xi_i \geq 0$이므로, $\| \mathrm{x}_i - \mathbf{a} \|^2 \geq \mathrm{R}^2$가 된다. 따라서, α_i가 0을 초과하는 벡터 x_i는 구체의 경계선 밖에 존재하는 서포트 벡터를 의미한다.

$$\alpha_i\left(\| x_i - a \|^2 - R^2 - \xi_i \right) = 0 \quad \forall i \qquad \text{(식 2.32)}$$

커널 함수의 도입을 통해 단순히 구형(球形)의 경계가 아닌 비선형의 경계를 구현할 수 있다. 식 2.31에 포함된 벡터 간의 내적 계산을 커널함수(kernel function)로 대체함으로써 데이터를 고차원의 커널 공간으로 맵핑할 수 있으며, 결과적으로 비선형의 유연한(flexible) 경계를 학습할 수 있게 된다. 커널 함수는 다항(polynomial) 커널, 가우시안(gaussian) 커널 등 다양한 커널 함수를 적용할 수 있으며, 본 연구에서는 가우시안 커널 함수를 적용하였다(식 2.33). 여기서, s는 가우시안 커널 함수의 파라미터를 의미한다.

$$K(x_i, x_j) = \exp\left(-\frac{\| x_i - x_j \|^2}{s^2}\right) \qquad \text{(식 2.33)}$$

가우시안 커널의 적용과 함께, SVDD 경계는 두 가지 파라미터 C와 s에 의해 결정된다. C는 초구체의 체적과 패널티 사이의 균형을 조절하는 파라미터로, 값이 증가함에 따라 패널티(ξ)에 대한 가중치가 증가하여 결과적으로는 초구체의 크기가 감소하게 되며, 반대로 값이 감소하면 초구체의 체적이 증가한다(그림 2-4). s는 SVDD 경계의 유연성(flexibility)을 결정하는 파라미터로, 값이 감소하면 SVDD의 경계는 좀 더 복잡하고 유연한 형태를 나타내며, 반대로 값이 증가하면 SVDD의 경계는 점차 구형에 가까워진다(그림 2-4).

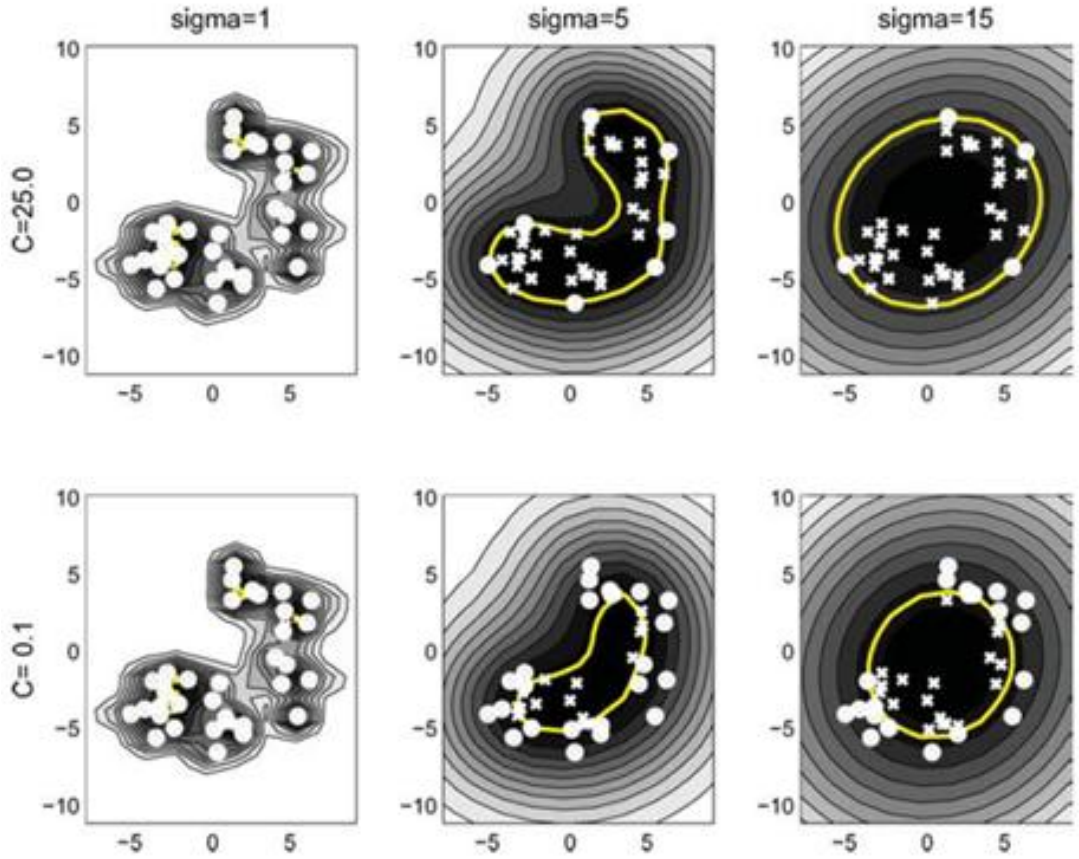


그림 2-4. 하이퍼 파라미터에 따른 SVDD 경계 형태[21)]

21) Tax, D. M. J. and Duin, R. P. W. (2004). Support vector data description. Machine Learning, 54(1), (pp. 45–66).

제 3 장 BEMS 데이터 분석 및 BNN 모델 제작

3.1 대상 건물 및 시스템

본 연구는 서울 소재의 연면적 32,600m^2, 지상 30층 및 지하 7층 규모 업무용 건물을 대상으로 수행되었다(그림 3-1). 해당 건물은 동일한 스펙의 압축식 냉동기 두 대가 빙축열 시스템의 냉열원으로 사용되고 있었으며, 각 냉동기는 정격 냉동용량 245.6USRT(United States refrigeration tons), 정격 소비전력 238.1kW, 정격 COP 4.81의 성능을 보유하고 있다. 냉동기는 비교적 가격이 저렴한 야간 전력을 사용하기 위해 오후 11시부터 익일 오전 3시까지 가동되며, 익일 사용될 얼음 캡슐을 생성한다.

그림 3-1. 대상 건물 및 압축식 냉동기

냉동기 데이터는 7월 10일부터 9월 30일까지 약 3달간 수집되었으며, 5분 간격으로 측정되었다. 소비전력을 기준으로 냉동기가 가동되지 않은 시점의 데이터는 분석에서 제외하였으며, 그 결과 총 23,904개(288개/일 ×83일) 데이터 중 2,672개 데이터가 분석에 사용되었다. 냉동기에서는 소

비전력(P), 브라인 입수온도($T_{brine,\ inlet}$), 브라인 출수온도($T_{brine,\ outlet}$), 브라인 유량($\dot{m}_{brine}$), 냉각수 입수온도($T_{cool,\ inlet}$), 냉각수 출수온도($T_{cool,\ outlet}$)가 수집된다(그림 3-2, 표 3-1). 냉동기 브라인 용액은 에틸렌글리콜 25% 수용액으로, 4℃에서 밀도는 1.042g/cm^3, 비열은 0.8818cal/gK이다. 브라인 용액의 특성치와 함께 브라인 입·출수 온도차, 브라인 유량, 전력량을 이용하여 시간대별 냉동기의 COP를 계산하고, 이를 분석에 포함하였다.

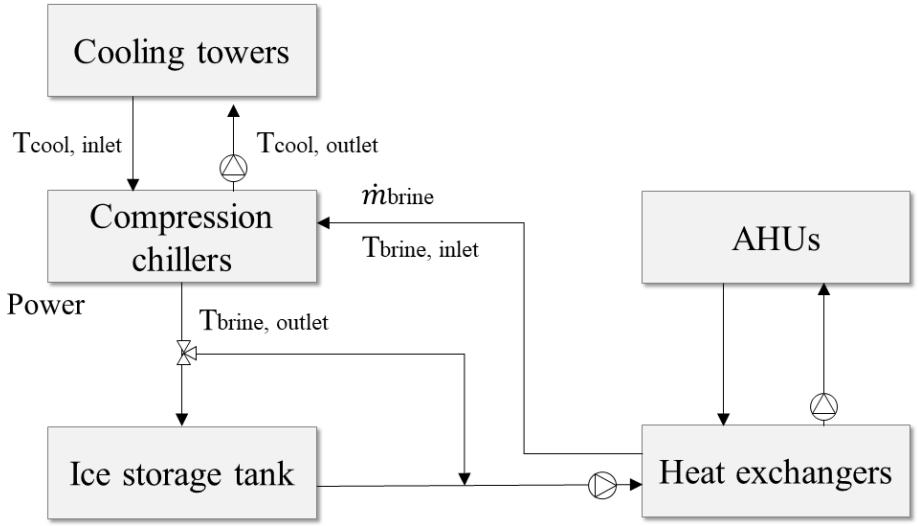


그림 3-2. HVAC 시스템 및 측정 데이터

표 3-1. 냉동기 데이터 수집 변수 목록

Measured data		unit
Brine inlet temperature from air-handling unit to chiller	$T_{brine,\ inlet}$	℃
Brine outlet temperature from chiller to air-handling unit	$T_{brine,\ outlet}$	℃
Cooling water inlet temperature from cooling tower to chiller	$T_{cool,\ inlet}$	℃
Cooling water outlet temperature from chiller to cooling tower	$T_{cool,\ outlet}$	℃
Brine's volumetric flow rate	$\dot{m}_{brine}$	m^3/h
Chiller power	P	kW
Coefficient of Performance (COP)	COP	-

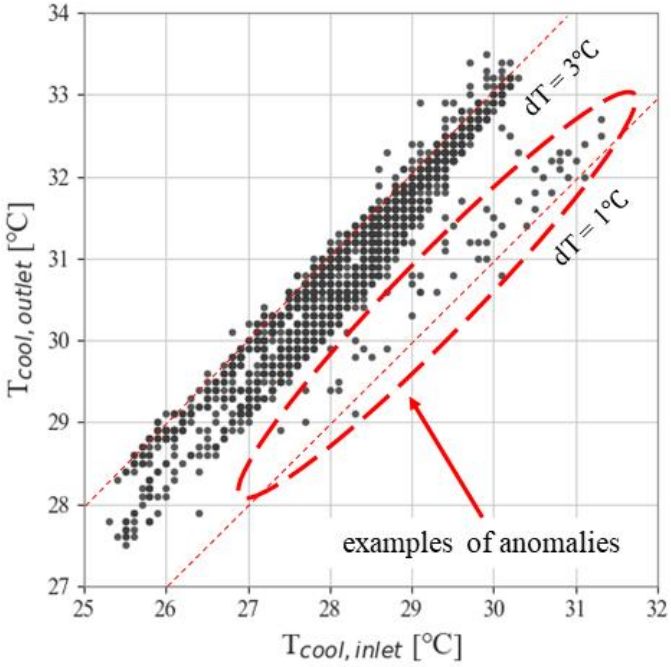
3.2 BEMS 데이터 분석

2장에서 언급한 바와 같이 BEMS 데이터 자체에 시스템의 이상 상태를 확인할 수 있는 레이블이 존재하지 않으므로, 데이터의 이상치 처리 결과에 대한 객관적인 검증이 불가능하다. 따라서 본 연구에서는 BEMS 데이터의 분석을 통해 시스템의 거동을 파악하고, 경험적 지식 및 판단에 기반하여 시스템의 비정상 거동으로 인해 발생한 것으로 추정되는 이상치 데이터들을 확인하였다.

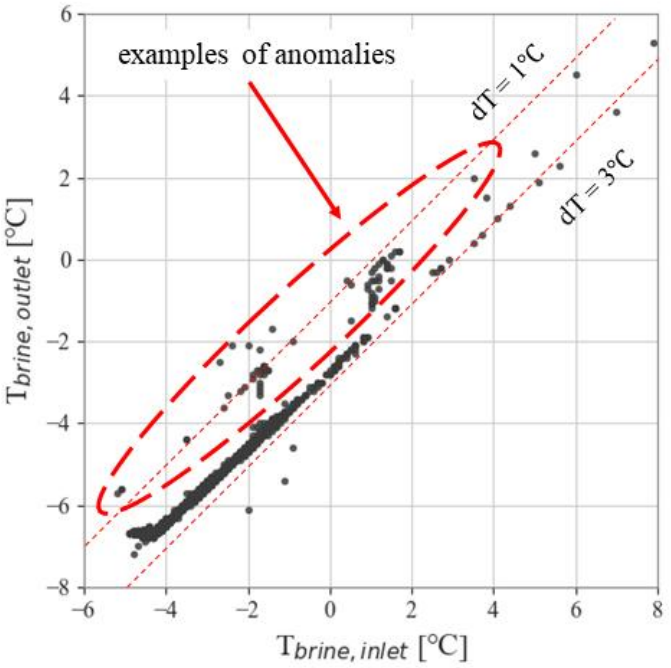
그림 3-3은 7월 10일부터 8월 26일까지의 냉동기 데이터를 나타낸 산점도로, 정상적인 데이터와 함께 일부 비정상 데이터들의 분포를 시각적으로 확인할 수 있다. 일반적으로 냉동기의 냉각수 입수온도($T_{cool,\ inlet}$)와 출수온도($T_{cool,\ outlet}$)의 온도 차는 2~3℃ 사이에서 운전되었지만, 그중 일부는 1℃까지 온도 차가 감소하는 것을 알 수 있었다(그림 3-3(a)). 마찬가지로, 브라인 입수온도($T_{brine,\ inlet}$)와 출수온도($T_{brine,\ outlet}$)의 온도 차 또한 3℃ 부근에서 주로 측정되는 반면, 일부 데이터가 1℃ 부근에서 측정되는 것을 알 수 있었다(그림 3-3(b)). 이러한 비정상 데이터들은 냉동기가 정상 상태에 도달하기 전 예열 단계, 혹은 가동 정지 시점에서 측정된 것으로 추정할 수 있다. 해당 비정상 데이터들은 그림 3-2에서 언급하였듯이, 개별 변수의 통계처리만으로는 식별이 어려운 다변수 이상치의 예시로 볼 수 있다. 예를 들어 냉각수 온도의 경우, 대부분의 비정상 데이터는 입수온도 범위인 25~31℃, 출수온도 범위인 27~34℃를 벗어나지 않는다(그림 3-3(a)). 또한, 냉동기 소비전력(Power)과 브라인 온도 차($T_{brine,\ inlet}$ - $T_{brine,\ outlet}$)는 양의 선형 상관관계를 가지고 있지만, 그림 3-3(c)의 붉은색 원으로 표시된 데이터는 이러한 상관관계를 따르지 않는다. 마찬가지로, 해당 냉동기의 정격 COP는 4.81이지만, 그림 3-3(d)의 붉은색 원으로 표시된 데이터들은 이를 크게 벗어난다.

상기 언급한 모든 비정상 데이터들이 센서의 오류나 냉동기의 오작동으로부터 측정된 것으로 보기는 어렵다. 하지만, 이러한 데이터들 역시 기계학습 모델을 제작할 때 사용된다면 모델의 성능이 저하될 수 있으며, 최적제어 수행 시에도 정확한 판단을 방해할 수 있다. 따라서, BEMS

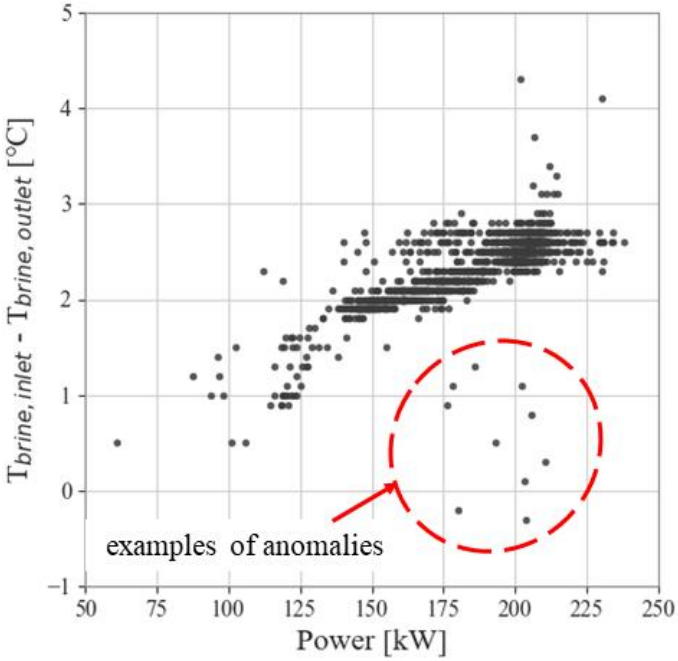
데이터를 이용한 분석 과정에서는 시뮬레이션 수행자의 직관적 개입 없이 데이터를 전처리하는 과정이 요구된다.



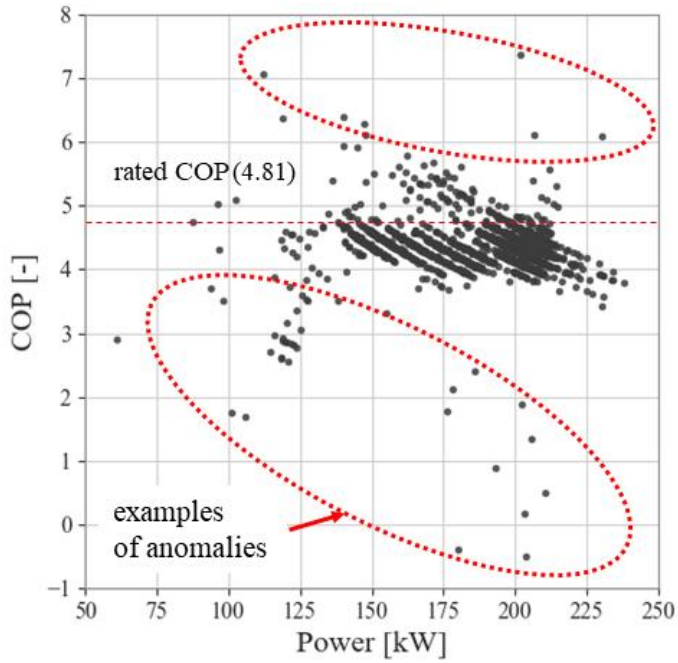
(a) 냉각수 입수 vs. 출수 온도



(b) 브라인 입수 vs. 출수 온도



(c) 소비전력 vs. 브라인 온도차



(d) 소비전력 vs. COP

그림 3-3. 냉동기 데이터 내 이상치 분포

3.3 이상치 검출 결과

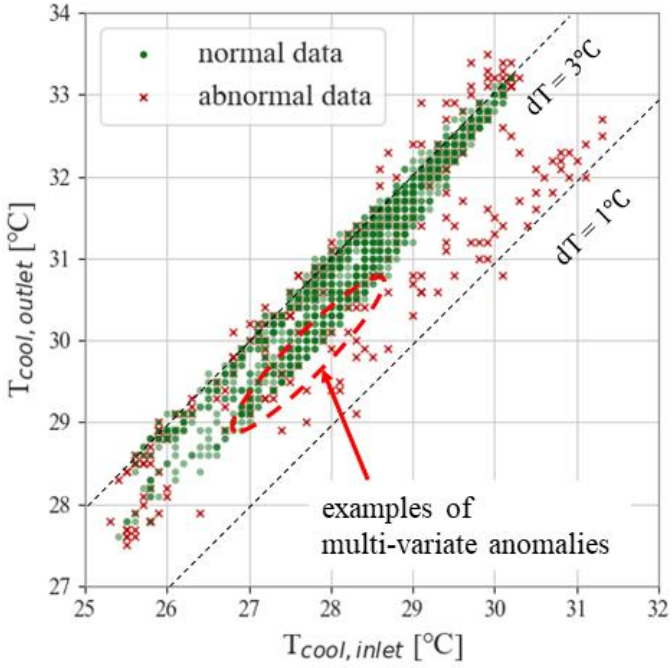
본 연구에서 이상치 식별 및 제거를 위해 사용한 SVDD 알고리즘은 Python 3.6 환경에서 구축했으며, 데이터 분석 라이브러리 Scikit-learn[22] 의 OneClassSVM(OCSVM) 모듈을 사용하였다. OCSVM 모듈에서는 두 가지 파라미터인 *nu*와 *gamma*를 통해 SVDD 경계를 결정한다. *nu*는 데이터의 에러 비율을 의미하는 값이며, *gamma*는 가우시안 커널의 파라미터이다. 데이터 내에 이상치가 약 15%가량 존재한다고 가정하여 *nu*는 0.15로 설정하였으며, *gamma*는 0.35로 설정하였다. 이는 시행착오(trial and error)를 통해 결정된 값이다. 그 결과, 총 1,918개 데이터 중 1,631개의 정상 데이터와 287개의 비정상 데이터로 분류되었다.

그림 3-4는 앞서 분석한 데이터(그림 3-3 참조)에 대해 SVDD를 이용해 이상치를 식별한 결과를 나타낸 산점도이다. 냉각수 입수온도와 출수온도 사이의 온도 차가 비정상적으로 1℃ 부근에 머물렀던 이상치(그림 3-3(a))가 SVDD를 통해 대부분 이상치로 식별되었음을 알 수 있다(그림 3-4(a)). 마찬가지로, 브라인 입·출수온도 차가 비정상적으로 1℃ 부근에 머물렀던 이상치(그림 3-3(b))도 SVDD가 이상치로 식별하였다(그림 3-4(b)). 또한, 소비전력과 브라인 입·출수온도 차 사이의 상관관계를 따르지 않았던 비정상 데이터(그림 3-3(c)), 정격 COP 4.81을 크게 벗어나는 비정상 데이터(그림 3-3(d)) 모두 SVDD가 이상치로 식별하였다(그림 3-3(c), (d)).

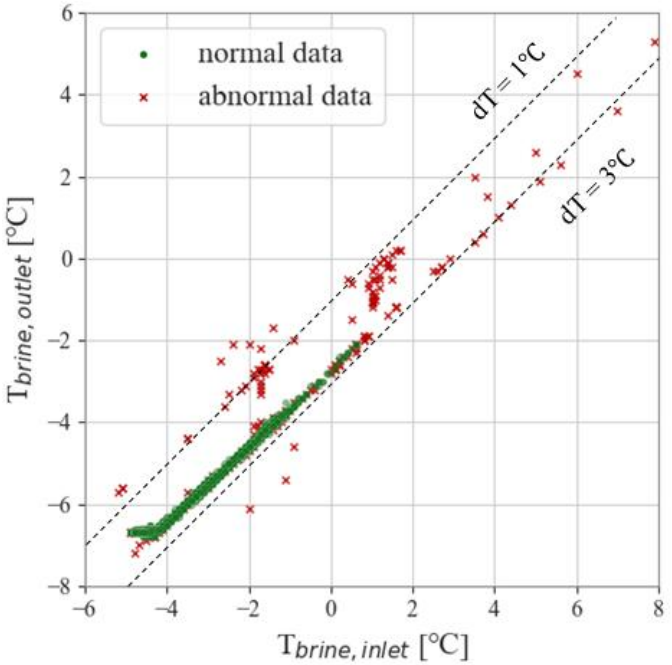
여기서 주목할 점은, 상기 언급한 비정상 데이터 외에도 이상치(그림 3-4의 붉은색 ×) 중 일부는 정상 데이터(녹색 •)의 범주 안에 존재하는 것처럼 보인다. 하지만, 이는 2차원 평면의 한계에 의한 착시일 뿐, n-차원 공간상에서는 해당 데이터들이 SVDD의 초구체 밖에 존재하기 때문에 이상치로 식별된 것이다. 이를 통해 BEMS 데이터 내 다차원 이상치의 존재를 확인할 수 있으며, SVDD가 다차원 이상치를 잘 식별했음을 알 수 있다.

---

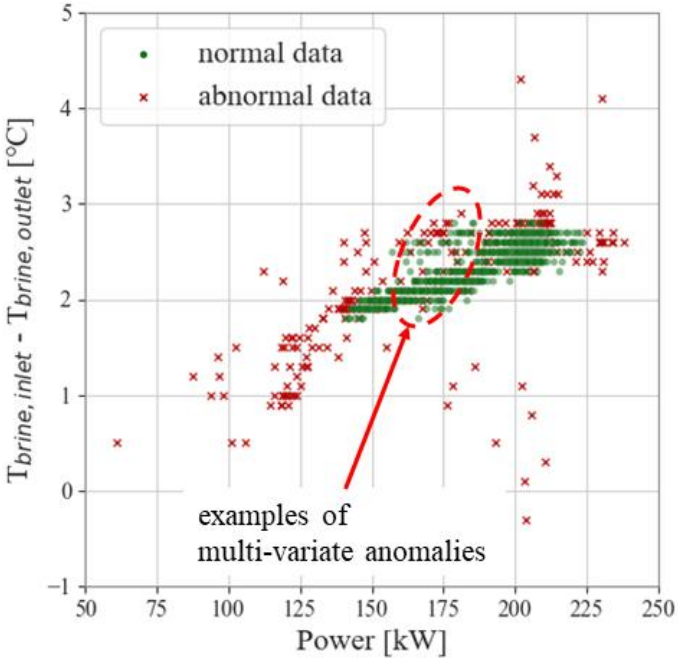
22) F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: Machine learning in Python, Journal of machine learning research, 12 (Oct) (2011) 2825-2830.



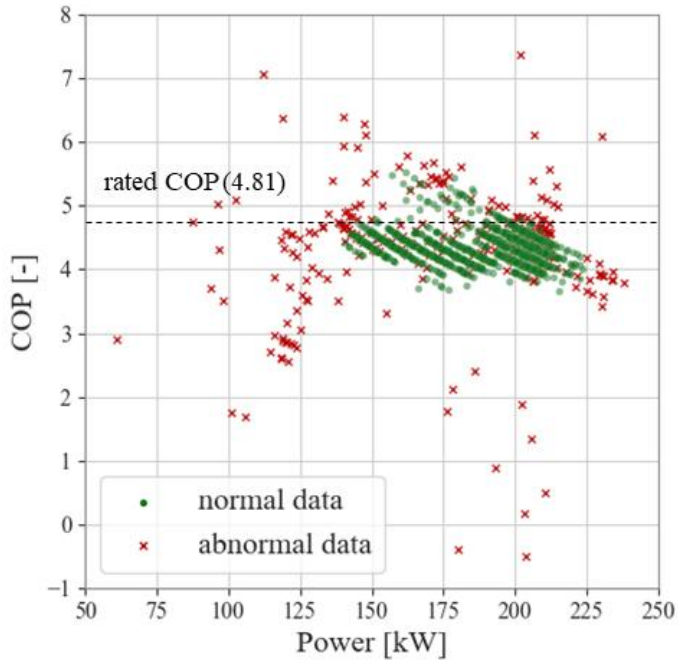
(a) 냉각수 입수 vs. 출수 온도



(b) 브라인 입수 vs. 출수 온도



(c) 소비전력 vs. 브라인 온도차



(d) 소비전력 vs. COP

그림 3-4. SVDD 이상치 검출 결과

제 4 장 베이지안 신경망 모델 제작 및 불확실성 분석

4.1 베이지안 신경망 모델 제작 및 검증

본 연구에서 베이지안 신경망은 냉동기의 운전 조건(입력변수)에 따른 COP(출력변수) 변화를 모사하도록 제작되었다. 냉동기 COP 예측을 위한 입력변수로는 브라인 입수온도, 냉각수 입수온도, 브라인 유량, 냉동기 전력을 선정하였다(표 4-1). 본 연구의 목적은 다양한 조건에서 제작되는 기계학습 모델의 불확실성을 정량적으로 확인하고, 훈련데이터의 양적, 질적 상태가 모델 불확실성에 미치는 영향을 확인하기 위함이다. 따라서, 훈련데이터의 기간과 이상치 제거 여부에 따라 훈련데이터를 총 4가지로 구분하여 베이지안 신경망 모델을 제작하였다(표 4-2). 첫 번째 신경망 모델의 훈련데이터는 7월 10일부터 7월 26일까지 약 16일간의 데이터를 사용하였으며, 이상치를 제거하지 않아 총 624개의 데이터가 학습에 사용되었다. 두 번째 신경망 모델의 훈련데이터는 첫 번째 모델과 동일한 16일간의 데이터를 사용하였으며, SVDD를 통해 15% 가량의 데이터를 이상 데이터로 제거하여 총 530개의 데이터가 학습에 사용되었다. 세 번째 모델의 훈련데이터는 기간을 7월 10일부터 8월 26일까지 늘린 약 48일간의 데이터를 사용하였으며, 데이터 이상치를 제거하지 않았으므로 제작된 4개의 모델 중 가장 많은 1,918개의 훈련데이터가 사용되었다. 마지막으로, 네 번째 모델은 48일간의 데이터 중 이상치를 제거하고 남은 1,631개의 훈련데이터로 학습하였다. 훈련데이터 외의 신경망 구조, 최적화 알고리즘, 기타 파라미터 등 모든 학습조건은 네 개 모델 모두 동일하게 설정하였으며, 신경망의 세부 설정은 표 4-3과 같다.

표 4-1. BNN 모델 입출력 변수

Input/Output	Measured data	unit
Inputs	Brine inlet temperature	℃
	Cooling water inlet temperature	℃
	Brine volumetric flow rate	m^3/h
	Power	kW
Output	COP	-

표 4-2. BNN 모델 훈련 및 검증 데이터

Model #	Train data period	Outliers	# of train data	Test data period
BNN #1	7.10~7.26	not removed	624	8.26-9.30
BNN #2	7.10~7.26	removed	530	
BNN #3	7.10~8.26	not removed	1,918	
BNN #4	7.10~8.26	removed	1,631	

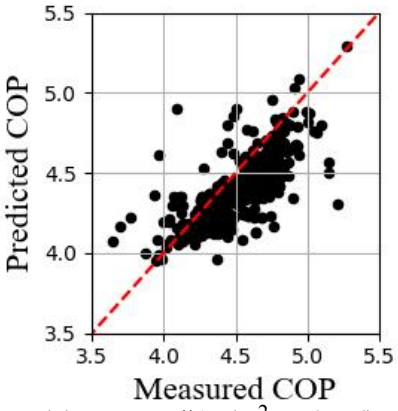
표 4-3. BNN 모델 파라미터

Parameter	Setting
epoch	500
activation function	leaky ReLU
number of hidden layers (nodes)	4 (7-13-25-9)
weight-decay	0.0001
mini-batch size	30
optimization algorithm	Rectified Adam
dropout rate	0.2
learning rate	0.01

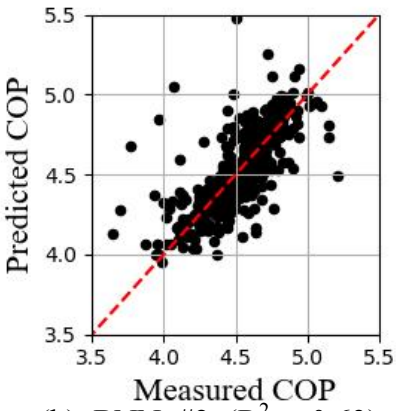
검증 기간 데이터(8월 26일부터 9월 30일까지)를 사용하여 4개의 BNN 모델의 예측 성능을 비교하였다. BNN 모델은 같은 입력값이라 할지라도, 무작위로 추출된 가중치에 따라 매번 다른 결과를 출력한다. 따라서, 해당 예측 성능 또한 3,000번의 반복 연산을 통해 예측값의 평균을 구한 뒤 측정값에 대해 예측 성능을 계산한 결과이다. 4개 BNN 모델의 예측 성능은 CVRMSE(Covariance of Root Mean Squared Error), MBE(Mean Bias Error), MAPE(Mean Absolute Percentage Error), 그리고 결정계수(coefficient of determination, R^2)의 측면에서 평가하였다. 참고로, ASHRAE Guideline 14(2014)에서는 시간별 예측 모델을 기준으로 CVRMSE 30%, MBE 10% 이하인 모델을 사용에 적합한 모델로 평가한다. 평가 결과, BNN #1의 MBE 만을 제외하고 모두 ASHRAE에서 권장하는 기준치를 만족하였으며, MAPE는 4개 모델 모두 5% 이하로 예측 성능이 우수했다(표 4-4). 결정계수의 경우, 0.5~0.63 사이의 준수한 성능을 보였다(표 4-4, 그림 4-1).

표 4-4. BNN 모델 예측 오차

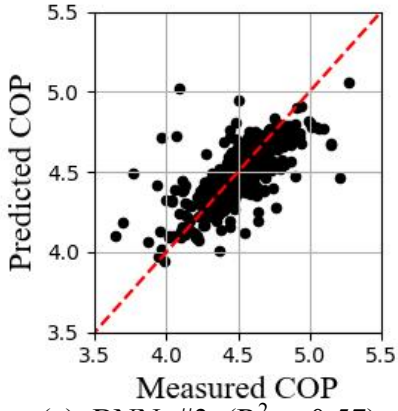
Model #	Test period (8.26 ~ 9.30)			
	CVRMSE (%)	MBE (%)	MAPE (%)	R^2
BNN #1	4.7	12.2	3.7	0.56
BNN #2	4.5	-1.8	3.0	0.63
BNN #3	3.7	2.0	2.6	0.57
BNN #4	4.0	5.0	2.8	0.5



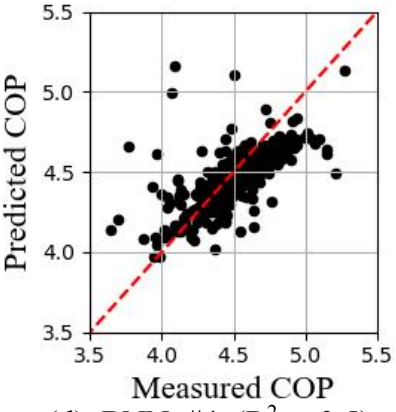
(a) BNN #1 (R^2 : 0.56)



(b) BNN #2 (R^2 : 0.63)



(c) BNN #3 (R^2 : 0.57)



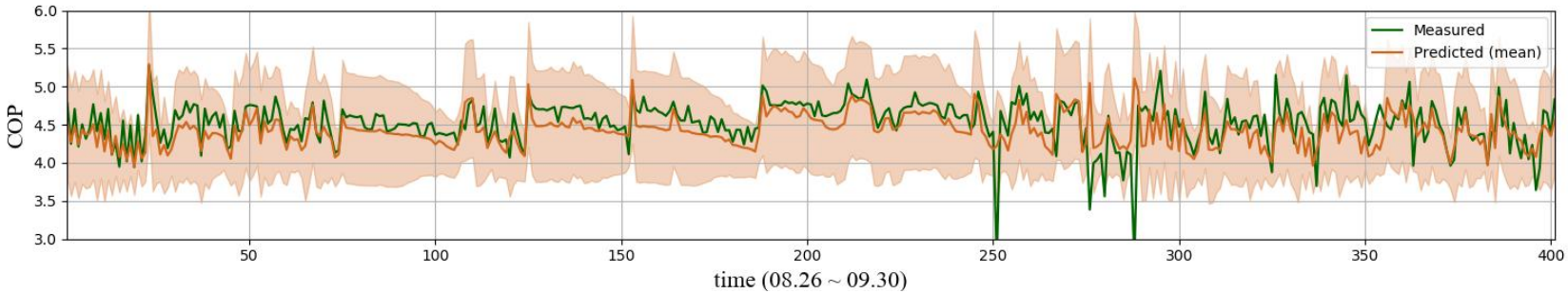
(d) BNN #4 (R^2 : 0.5)

그림 4-1. 검증 기간 BNN 모델 예측 성능(측정 COP vs 예측 COP)

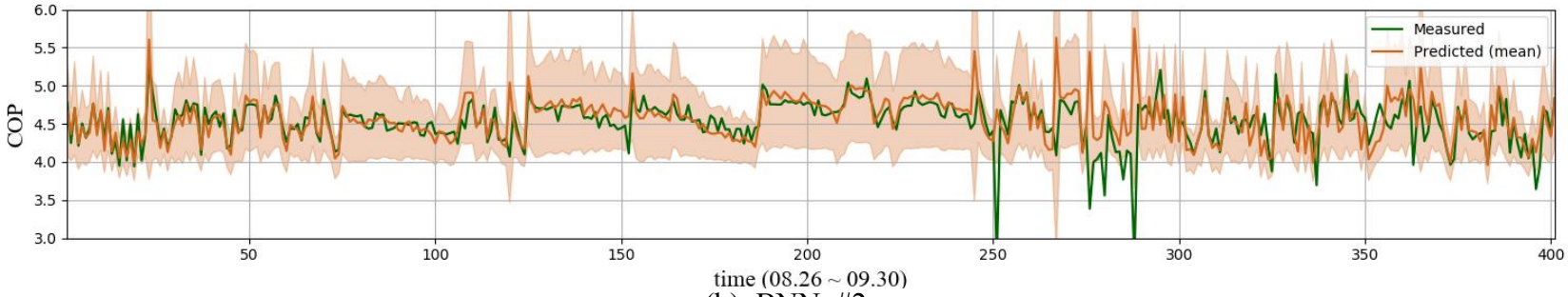
4.2 모델 불확실성 분석

4.1절에서는 BEMS 데이터를 통해 4개의 BNN 모델을 제작하고, 검증 기간 데이터를 통해 모델의 예측 성능을 비교하였다. 결과적으로, 제작된 4개의 BNN 모델 모두 예측 성능이 우수하였다. 하지만, 해당 예측 성능에 대한 평가는 3,000번의 반복 연산한 결과의 평균을 통해 계산된 결과로, 모델의 평균적인 예측 성능을 의미할 뿐, 매번 동일한 성능을 가진다고 볼 수 없다. 예측 성능이 우수한 모델이라 할지라도 모델 혹은 데이터가 불확실성을 발생시키는 요인을 내재하고 있다면, 모델 학습 과정에서 매번 전혀 다른 결과를 출력할 수 있으며, 그로 인하여 예측 성능이 저하될 수 있다.

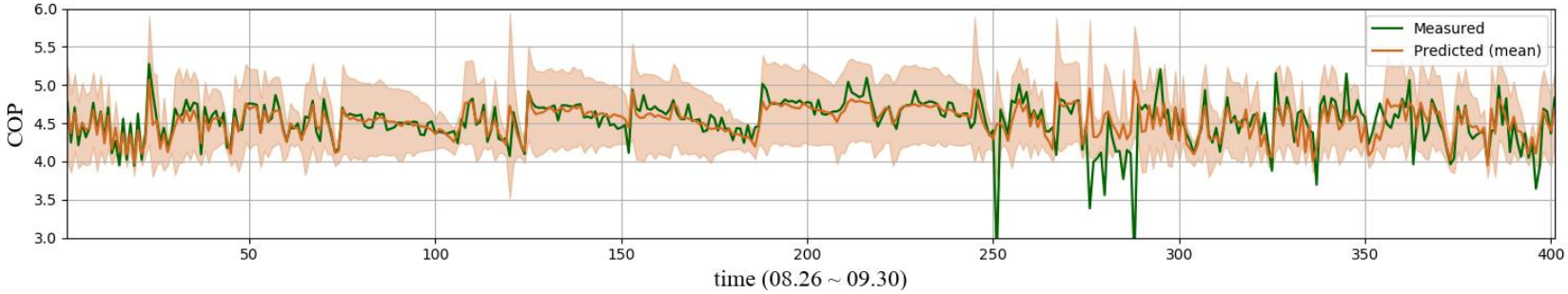
본 절에서는 제작된 BNN 모델에 내재된 불확실성에 대해 분석하였다. 먼저, 그림 4-2는 검증 기간(8월 26일-9월 30일, 표 4-2) 데이터에 대하여 BNN 모델의 예측값(주황색 실선)과 측정값(녹색 실선)을 나타낸 그림이다. 그림 4-2에서 주황색으로 표시된 영역은 모델의 불확실성 범위를 나타내는 것으로, 모델 예측의 68%(±1 표준편차) 신뢰구간을 의미한다. 여기서, 표준편차는 식 2.24를 통해 계산된 전체(인식론적 + 내재적) 불확실성의 제곱근으로, 두 불확실성의 합이 모델 출력(COP)에 대한 분산을 의미하므로, COP 단위로 환산하기 위해 제곱근을 취했다. 일부 구간(250~280 부근)을 제외한 대부분 시간에 대해, 4개 BNN 모델 모두 예측 평균(주황색 실선)이 측정값(녹색 실선)의 거동을 잘 모사하였다. 하지만, 모델 불확실성의 측면에서는 4개의 모델이 규모의 차이를 보였다. BNN #1의 경우(그림 4-2(a)) 예측 불확실성이 예측값(평균)을 기준으로 ±1 내외로 발생한 반면, BNN #4(그림 4-2(d))의 불확실성 범위는 ±0.5 이내로 BNN #1에 비해 비교적 감소함을 확인할 수 있었다.



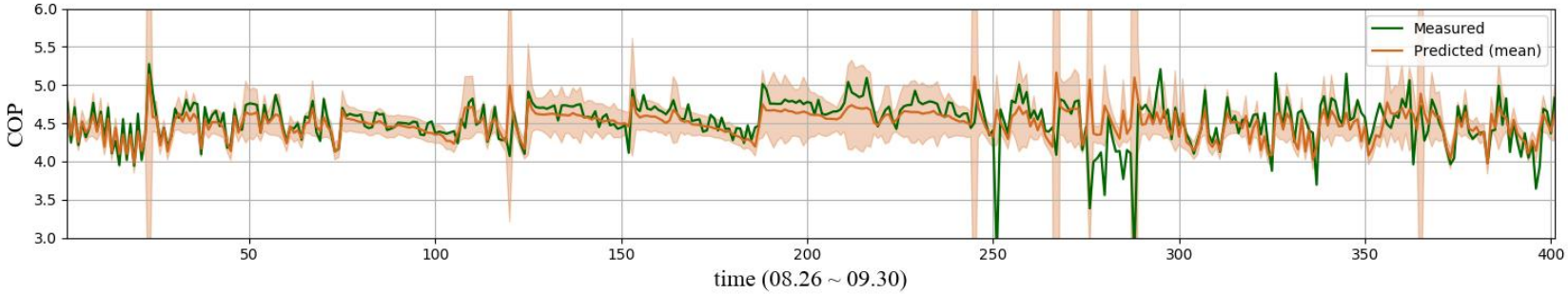
(a) BNN #1



(b) BNN #2



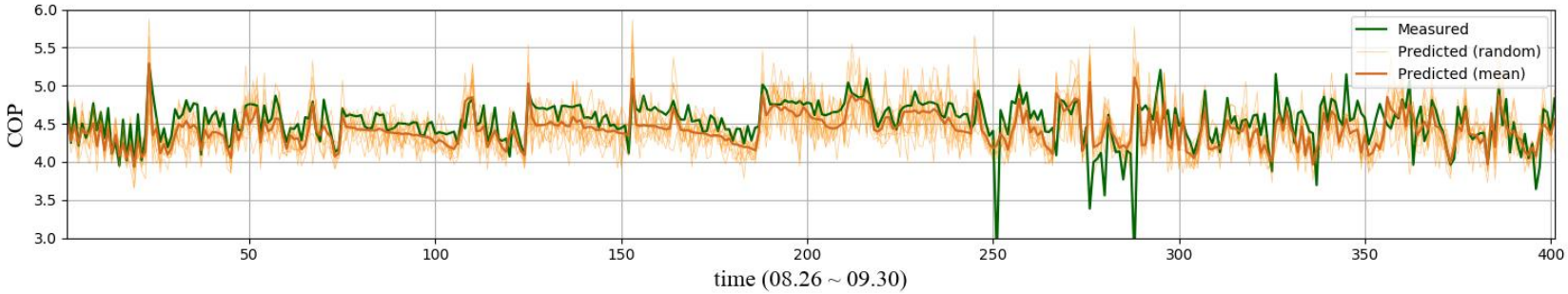
(c) BNN #3



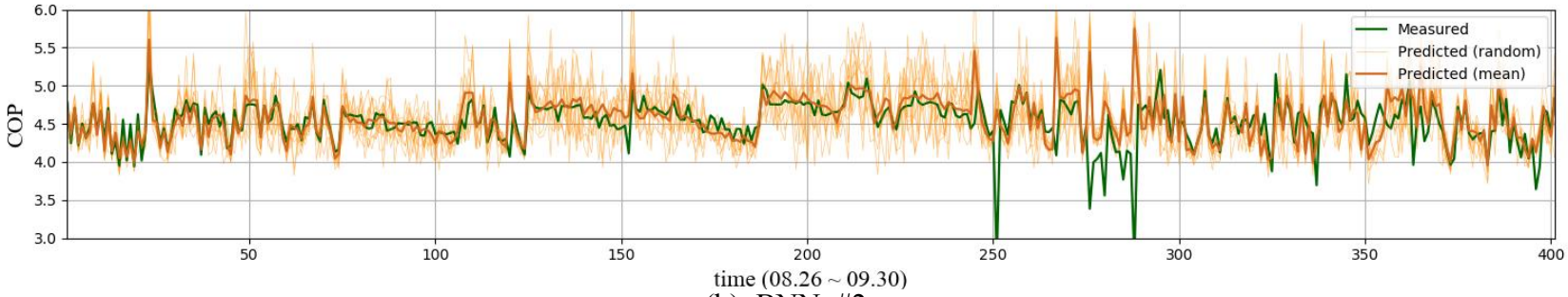
(d) BNN #4

그림 4-2. 모델 예측 결과 및 불확실성 범위

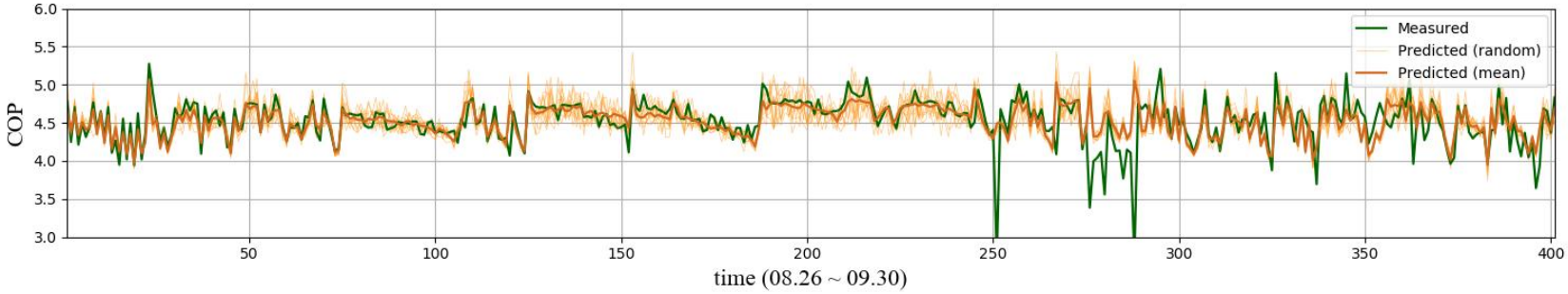
그림 4-3은 모델의 예측값, 측정값과 함께 무작위로 추출된 10개 가중치 집합에 의해 계산된 예측 결과(옅은 실선)를 표현한 그림이다. 예측 불확실성이 다른 모델에 비해 비교적 큰 BNN #1은 무작위로 출력되는 예측 결과 또한 변동이 크다(그림 4-3(a)). 이러한 모델은 가중치의 선택에 따라 모델의 예측 성능이 크게 좌우될 수 있으며, 모델의 학습 과정에서 평균적인 예측 성능에 미치지 못하는 신경망 모델을 얻게 될 위험성이 존재하게 된다. 반면에 불확실성의 규모가 비교적 작은 BNN #4의 경우, 무작위로 출력되는 예측값의 변동 또한 BNN #1에 비해 적으며(그림 4-3(d)), 이는 가중치의 선택과 상관없이 안정적인 예측 성능을 확보할 수 있음을 의미한다.



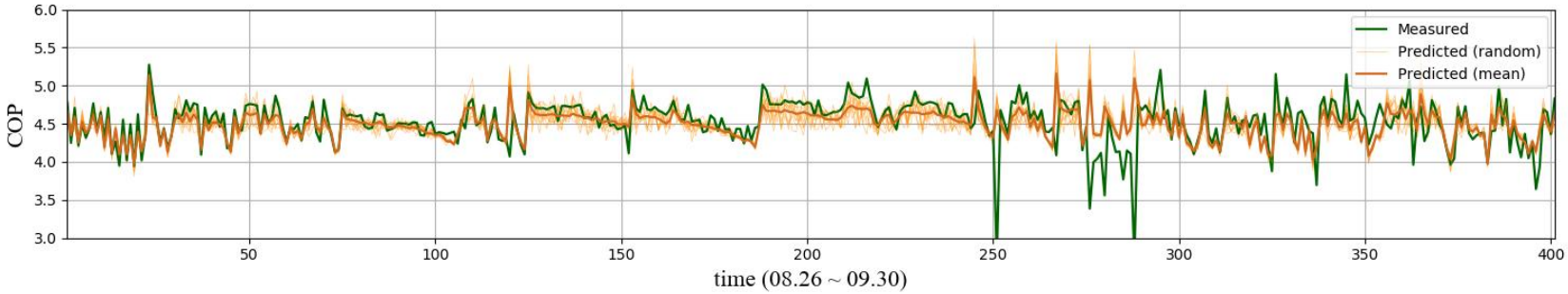
(a) BNN #1



(b) BNN #2



(c) BNN #3

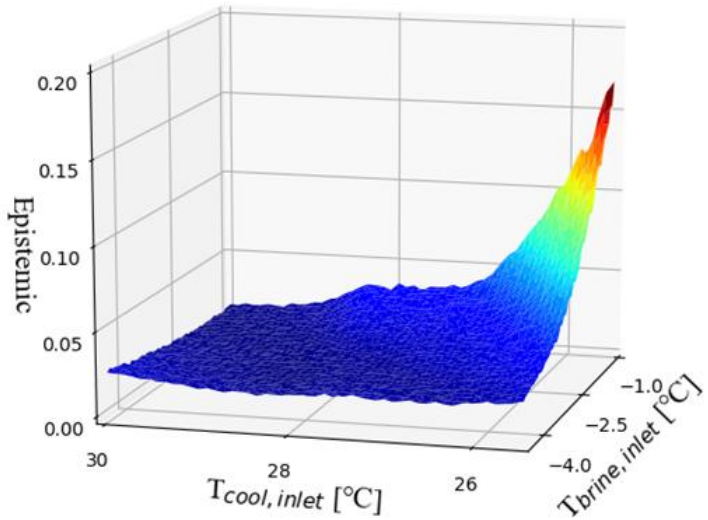


(d) BNN #4

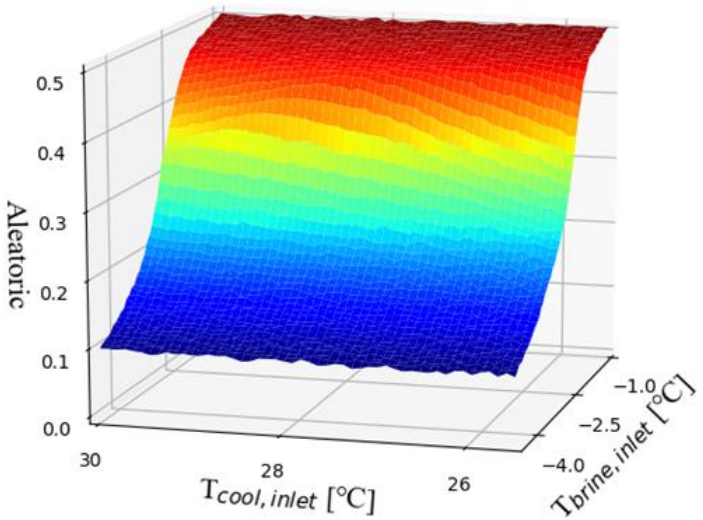
그림 4-3. 무작위 가중치에 의한 모델 예측 결과

그림 4-4 ~ 4-7는 냉동기 전력과 브라인 유량을 최빈값으로 고정하고 (전력: 200kW, 유량: 323m^3/h), 브라인 입수온도($T_{brine,\ inlet}$), 냉각수 입수온도($T_{cool,\ inlet}$)를 변화시켰을 때, COP 예측 결과에 대한 불확실성을 정량화한 결과이다. 이때, 브라인 입수온도와 냉각수 입수온도는 냉동기 BEMS 데이터를 기준으로 일반적인 가동 범위(브라인 입수온도는 -4.5℃~-1.0℃, 냉각수 입수온도는 25.5℃~30℃ 범위) 내에서 변화시켰다. BNN 모델의 인식론적 불확실성은 4개 모두 브라인 입수온도가 -1℃, 냉각수 입수온도가 25.5℃인 부근에서 최대로 발생하였으며, 이는 해당 부근에서 훈련데이터가 부족하여 나타난 결과로 추정된다. 내재적 불확실성은 브라인 입수온도가 낮은 -4℃ 부근에서 최소로 발생하였으며, 온도가 -1℃ 부근에 가까워질수록 불확실성이 증가하였다. BNN #1과 #2의 인식론적 불확실성은 최대 0.17 전후로 모델 간의 큰 차이를 보이지 않았다(그림 4-4(a), 4-5(a)). 이는 두 모델의 훈련데이터 기간이 동일하며, 훈련데이터의 개수에도 큰 차이를 보이지 않았기 때문이다(표 4-2참조). 반면, 내재적 불확실성의 비교에서는 BNN #2가 #1보다 적은 불확실성을 나타냈다(그림 4-4(b), 4-5(b)). 이는 SVDD를 통해 훈련데이터 내 존재하는 이상치를 제거함으로써 내재적 불확실성이 감소한 것으로 추정할 수 있다. 전체 불확실성 또한, BNN #1은 최대 0.8까지 발생했지만, BNN #2는 0.7 이내로 감소하였다. 결과적으로, BNN #1과 #2는 비슷한 예측 성능(CVRMSE 기준 BNN #1: 4.7%, BNN #2: 4.5%)을 가졌음에도 불구하고 (표 3-5), 예측 불확실성의 측면에서는 차이를 보임을 알 수 있었다. BNN #1과 #3의 경우, 두 모델 모두 훈련데이터의 이상치를 제거하지 않은 원본 데이터를 학습하였지만, BNN #3이 #1에 비해 3배가량 많은 훈련데이터를 사용하였다(624개 → 1,918개, 표 4-2 참조). 그 결과, BNN #1에서 0.17까지 발생했던 인식론적 불확실성은 BNN #3에서 0.05 이내로 감소했다(그림 4-4(a), 4-6(a)). 마지막으로 BNN#1과 #4의 비교 결과, 훈련데이터의 기간이 증가하고 훈련데이터 내의 이상치를 제거함으로써 모델의 인식론적 불확실성과 내재적 불확실성이 모두 감소했으며, 전체 불확실성이 최대 0.8에서 0.5 이내로 감소하였다(그림 4-4(c), 4-7(c)). 결

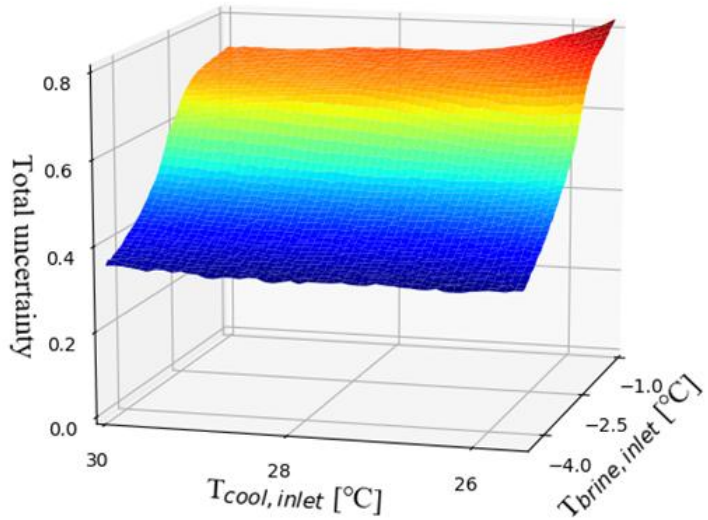
과적으로, 불확실성이 비교적 큰 BNN #1의 경우, 예측 결과에 대한 불확실성 범위가 넓은 반면(그림 4-4(d)), BNN #4는 같은 입력지점에 대해 불확실성 범위가 현저히 작아지는 것을 확인할 수 있다(그림 4-7(d)).



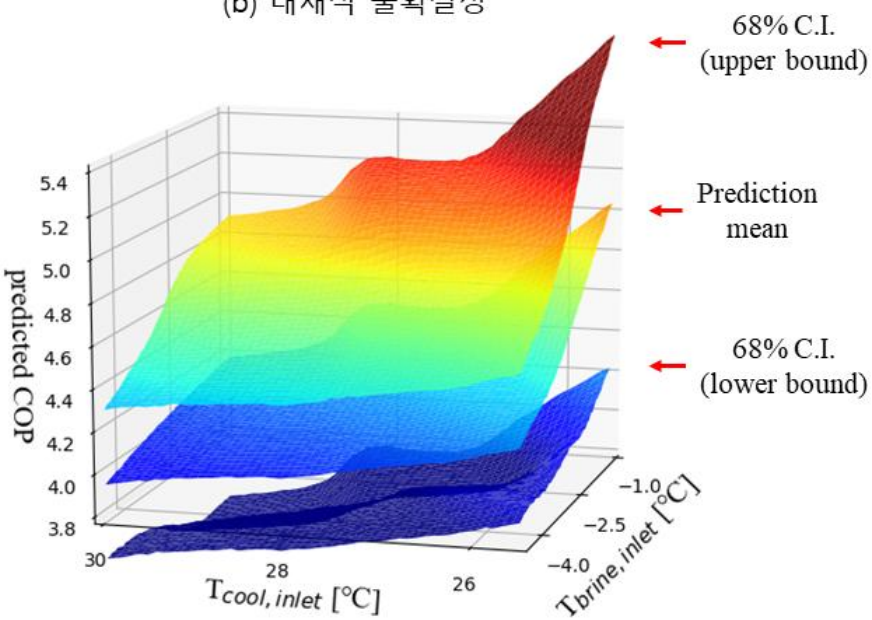
(a) 인식론적 불확실성



(b) 내재적 불확실성

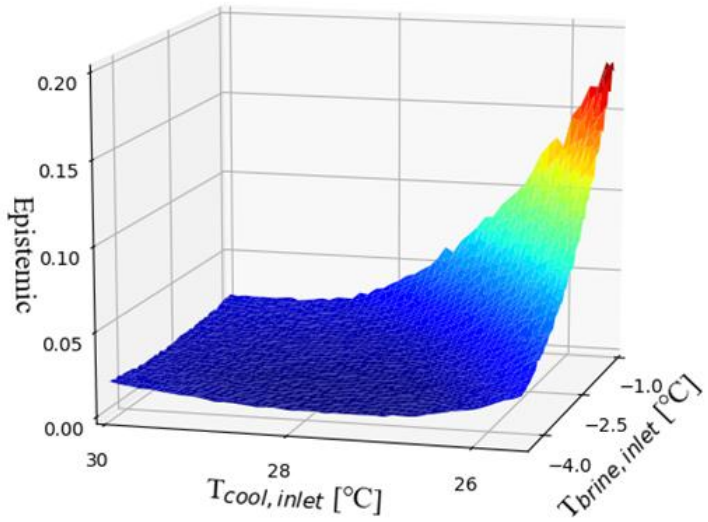


(c) 전체 불확실성

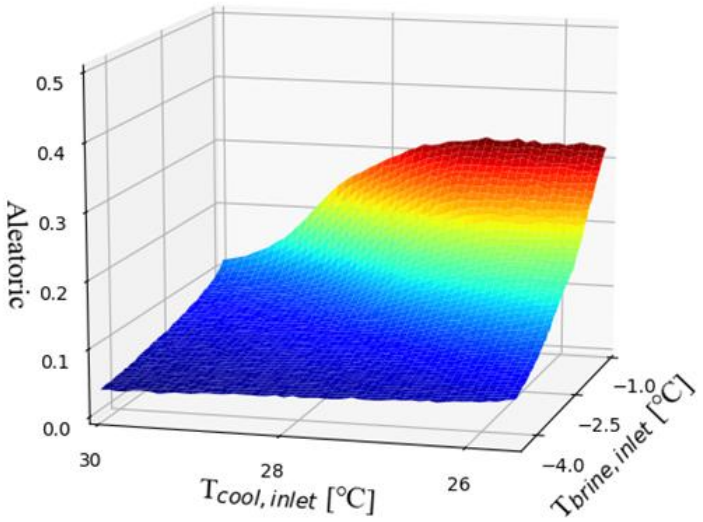


(d) 불확실성 범위

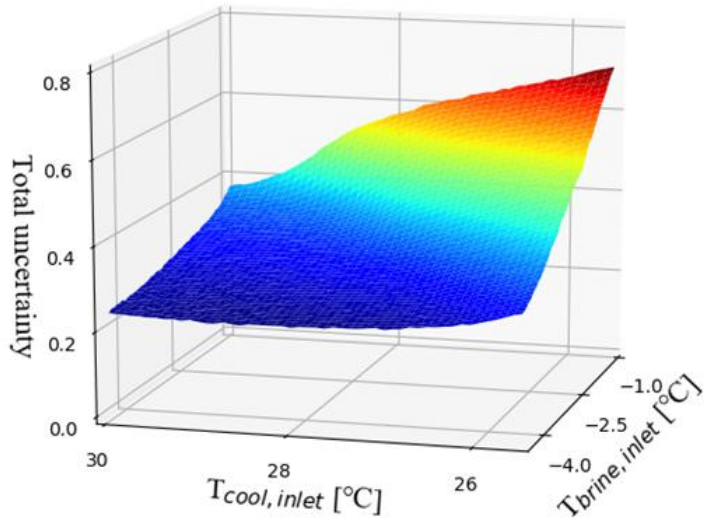
그림 4-4. 모델 불확실성 정량화 결과 (BNN #1)



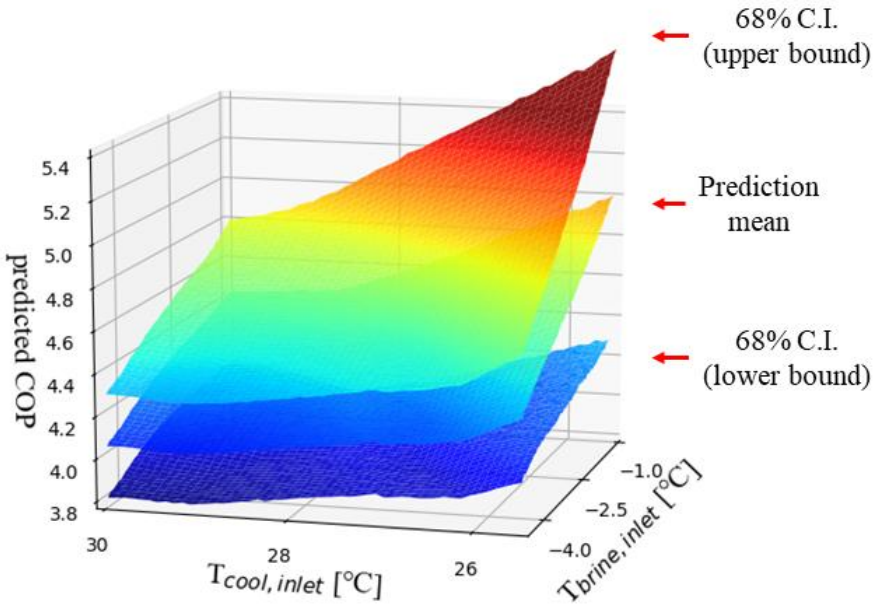
(a) 인식론적 불확실성



(b) 내재적 불확실성

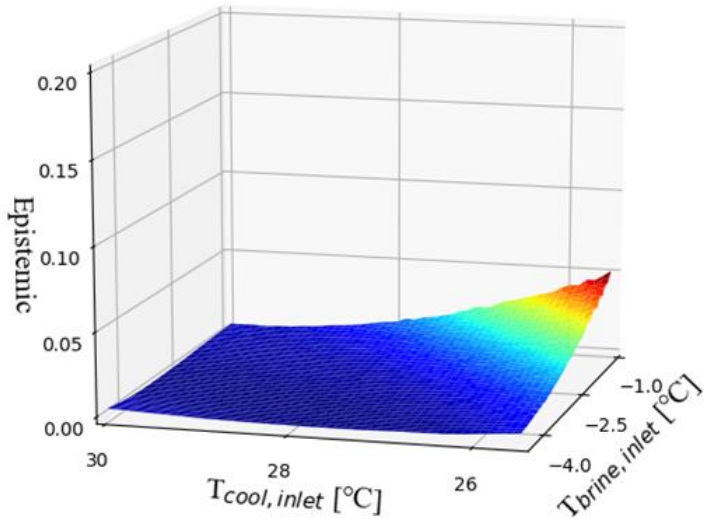


(c) 전체 불확실성

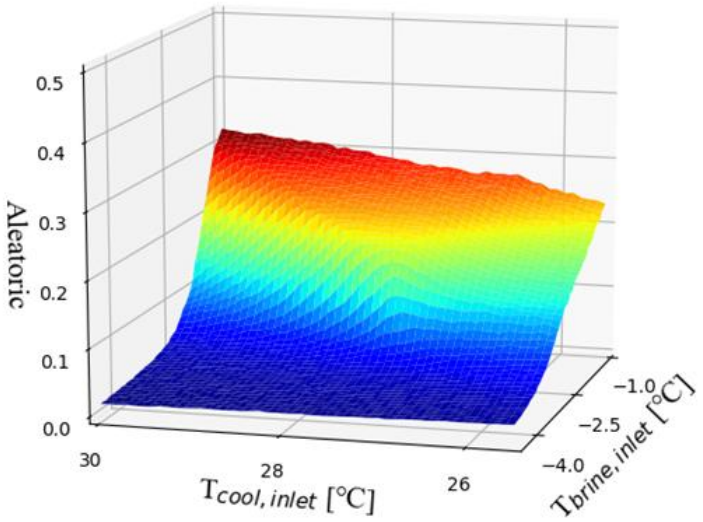


(d) 불확실성 범위

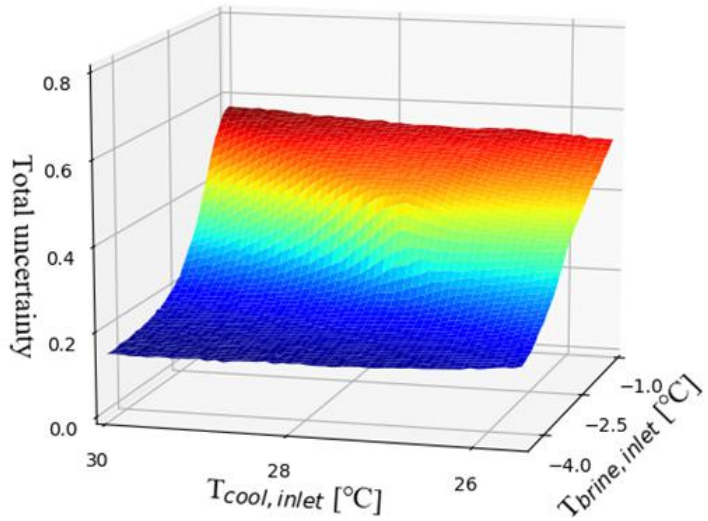
그림 4-5. 모델 불확실성 정량화 결과 (BNN #2)



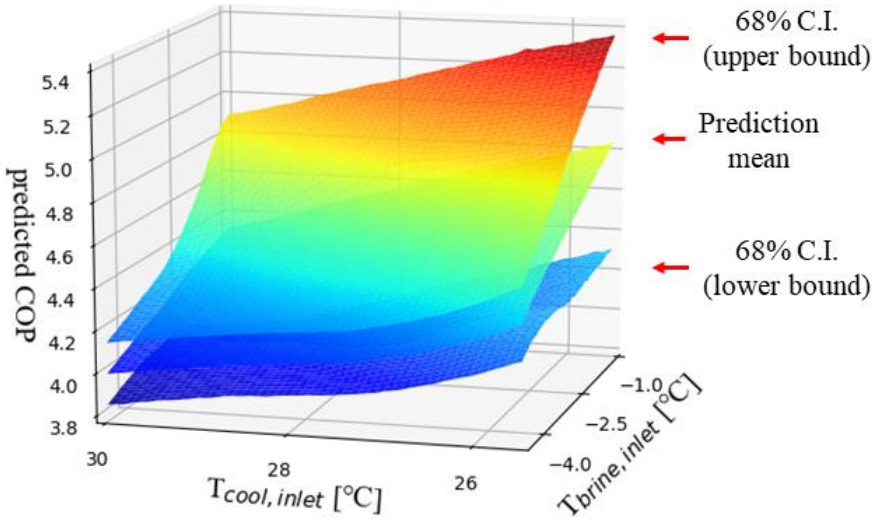
(a) 인식론적 불확실성



(b) 내재적 불확실성

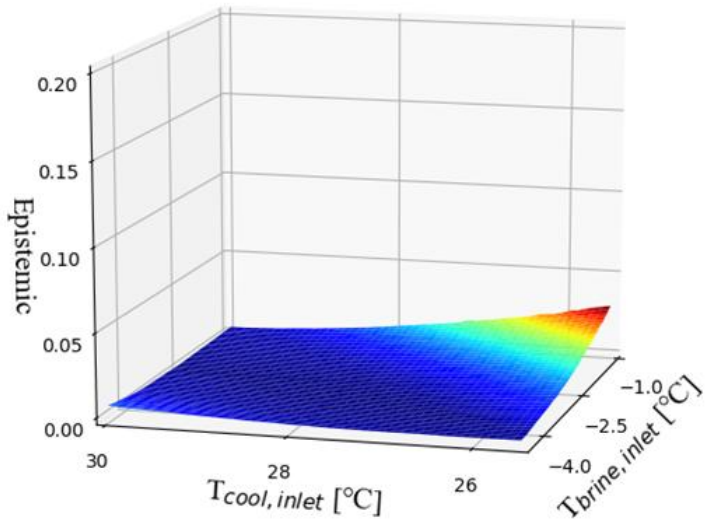


(c) 전체 불확실성

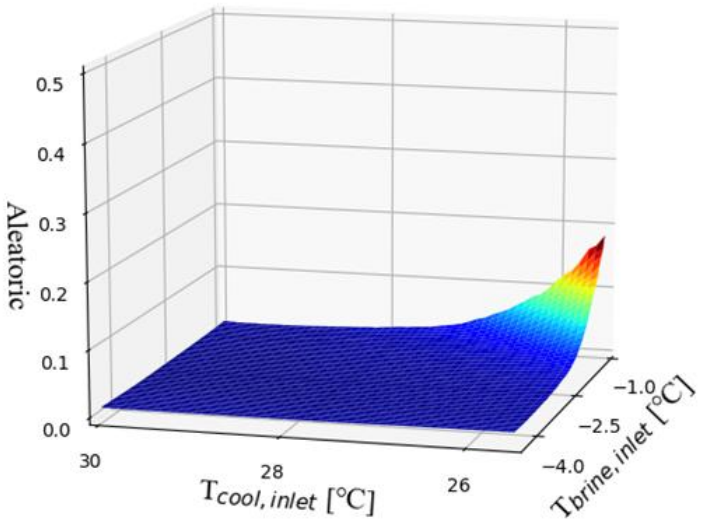


(d) 불확실성 범위

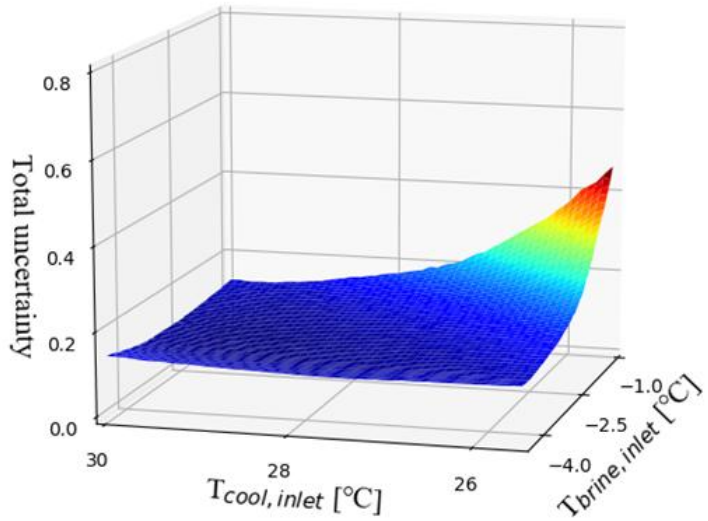
그림 4-6. 모델 불확실성 정량화 결과 (BNN #3)



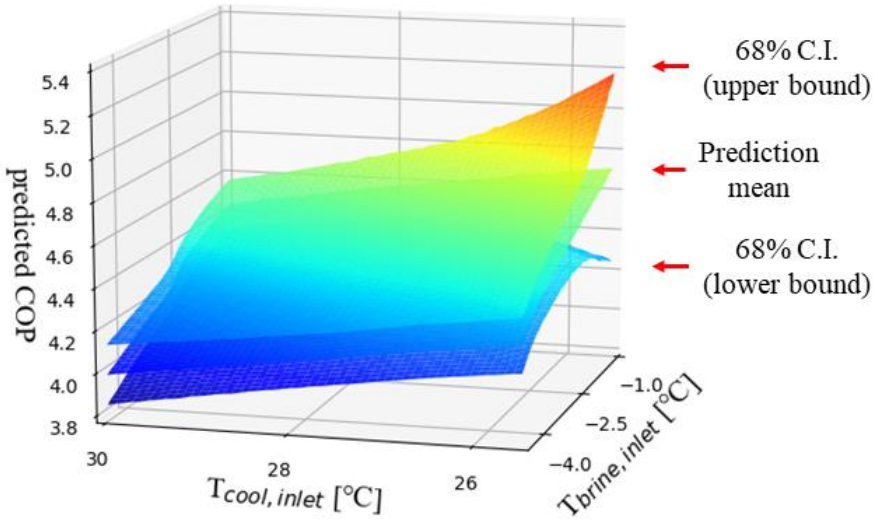
(a) 인식론적 불확실성



(b) 내재적 불확실성



(c) 전체 불확실성



(d) 불확실성 범위

그림 4-7. 모델 불확실성 정량화 결과 (BNN #4)

표 4-5은 검증 기간(8월 26일-9월 30일, 표 4-2)에 대한 네 가지 모델의 예측 오차(표 4-4 참조)와 그림 4-4 ~ 4-7의 불확실성 결과를 요약한 것이다. 4개 모델의 비교 결과, 훈련데이터의 기간에 따라 인식론적 불확실성($U_{epistemic}$)의 평균은 0.028에서 0.004까지 감소하였으며, 최댓값은 0.168에서 0.029까지 감소하였다. 또한, 훈련데이터의 이상치 검출 여부에 따라 내재적 불확실성($U_{aleatoric}$)의 평균은 0.306에서 0.017까지 감소하였고, 최댓값은 0.504에서 0.183까지 감소하였다. 두 불확실성의 합을 COP 단위로 환산하였을 때(표 4-5의 Total에 해당), BNN #1은 출력값의 불확실성이 평균 0.558, 최대 0.818까지 발생했다. 이를 정격 COP인 4.81과 비교하면, 평균 11.6% (=0.558/4.81×100)에서 최대 17.0% (=0.818/4.81×100)의 예측 불확실성이 발생함을 알 수 있다. 반면, 훈련데이터의 양적, 질적 품질이 향상됨에 따라 (훈련데이터 수의 증가 및 이상치 제거, 표 4-2 참조), BNN #4의 불확실성은 평균 0.137, 최대 0.460까지 감소했으며, 평균 2.8% (=0.137/4.81×100), 최대 9.6% (=0.460/4.81×100)의 불확실성을 보였다. 이처럼 신경망 모델의 학습에 사용된 훈련데이터의 상태에 따라서 모델의 불확실성에 차이를 보이며, 데이터의 수집 및 이상치 처리 과정을 통해 인식론적 불확실성 및 내재적 불확실성을 줄일 수 있음을 확인하였다. 또한, 불확실성을 정량화함으로써 훈련데이터의 품질에 대한 평가 수단으로써 활용될 수 있음을 확인하였다.

표 4-5. BNN 모델 예측 오차 및 불확실성

Model #	(CV) RMSE (%)	$U_{epistemic}$ (a)		$U_{aleatoric}$ (b)		Total ($\sqrt{a+b}$)	
		mean	max	mean	max	mean	max
BNN #1	4.7	0.027	0.168	0.306	0.504	0.558	0.818
BNN #2	4.5	0.028	0.180	0.113	0.326	0.357	0.707
BNN #3	3.7	0.007	0.050	0.122	0.320	0.326	0.569
BNN #4	4.0	0.004	0.029	0.017	0.183	0.137	0.460

제 5 장 결론

베이지안 신경망은 신경망 가중치에 대한 확률적 해석을 통해, 기계학습 모델에 내재된 불확실성을 정량적으로 평가할 수 있다는 장점이 있다. 본 연구에서는 베이지안 신경망을 이용하여, 기계학습 모델에 내재된 불확실성을 정략적으로 평가하고 훈련데이터의 양적, 질적 품질에 의해 변화하는 두 가지 불확실성(인식론적 및 내재적 불확실성)을 통계 수치와 시각화를 통해 확인하였다.

본 분석을 위해 실제 업무용 건물에 설치된 BEMS에서 수집된 냉동기 데이터를 대상으로, 냉동기의 가동 조건(입수온도, 유량, 전력)에 따른 COP 변화를 예측하는 베이지안 신경망 모델을 제작하였다. 해당 모델은 본 논문의 목적인 기계학습 모델의 불확실성을 정량화하기 위해 제작된 예시 모델로서, 모델 학습을 위한 훈련데이터는 수집 기간과 이상치 처리 여부에 따라 4개 세트로 구성하고 각각의 훈련데이터 세트를 통해 총 4개의 베이지안 신경망 모델을 제작하였다. 제작된 모델의 예측 성능을 검증 기간 데이터에 대해 CVRMSE, MBE, MAPE, R^2 등으로 평가하였을 때, 4개의 모델 모두 준수한 예측 성능을 보였다. 하지만, 해당 예측 성능과는 관계없이 모델에 내재된 불확실성의 크기는 모두 달랐다. 불확실성이 큰 모델의 경우 신경망 가중치의 무작위 선정에 따라 예측 결과의 변동 또한 컸으며, 반대로 불확실성이 작은 모델은 가중치의 무작위 선정에도 매번 비슷한 예측 결과를 나타내었다. 본 분석에서의 예측 성능은 3,000번의 반복 연산에 대한 평균적인 예측 성능을 의미할 뿐, 해당 모델이 항상 같은 성능을 보인다고 단정할 수 없다. 같은 구조의 신경망 모델일지라도 하이퍼 파라미터 선택, 초깃값 설정, 최적화 과정에서의 확률적 특성 등에 의해 매번 다른 가중치를 학습할 수 있으며, 이로 인해 예측 결과 또한 달라질 수 있다. 만일, 신경망 모델이 높은 불확실성을 내재하고 있다면 가중치 선택에 따라 모델의 예측 성능이 크게 좌우될 수 있으며, 높은 불확실성을 내재한 기계학습 모델을 최적제어나 의사결

정의 수단으로 사용하는 것은 안정성에 대한 위험이 따를 수 있다. 본 분석을 통해, 검증 기간 데이터에 대한 모델의 예측 성능이 우수할지라도 내재된 불확실성에 따라 모델 예측의 안정성이 떨어질 수 있음을 확인할 수 있었으며, 기계학습 모델의 예측 성능을 평가할 때는 검증데이터에 대한 예측 오차뿐만 아니라 예측 결과에 대한 불확실성을 정량적으로 분석하는 것이 필요하다는 것을 알 수 있었다.

또한, 훈련데이터의 양적, 질적 품질에 따른 인식론적 및 내재적 불확실성의 변화를 분석한 결과, 훈련데이터의 기간이 증가할수록 신경망 모델의 인식론적 불확실성이 감소함을 확인할 수 있었다(BNN #1→#3 or BNN #2→#4). 반면에 SVDD를 통해 훈련데이터 내의 이상치를 제거할 경우 내재적 불확실성이 감소하는 것을 확인하였다(BNN #1→#2 or BNN #3→#4). 결과적으로, 정격 COP 4.81에 대한 전체(인식론적 및 내재적) 불확실성 크기의 비율이 BNN #1의 경우 11.6%였으나, 훈련데이터의 추가 및 이상치 검출을 통해 2.8%(BNN #4)까지 감소함을 확인하였다.

기계학습 모델의 불확실성을 정량화하는 것은 모델 예측 결과에 대한 신뢰도를 평가할 수 있는 지표를 제공할 뿐만 아니라, 해당 신뢰도를 통해 최적제어의 목적함수 또는 의사결정의 수단으로써 사용될 수 있다. 또한, 신경망 모델의 불확실성을 인식론적 및 내재적 불확실성으로 분리함으로써 분석가에게 추가 정보를 제공할 수 있다. 예를 들어, 인식론적 불확실성을 통해 훈련데이터와 검증데이터의 상이한 수준을 평가할 수 있으며, 이는 추가 데이터 수집에 대한 필요성을 판단할 수 있게 해준다. 또한, 인식론적 불확실성은 신경망의 구조에 대한 평가 지표를 제공하여 신경망의 하이퍼 파라미터를 최적화하기 위해 사용될 수 있다. 반면, 내재적 불확실성은 훈련데이터 내 이상치 및 노이즈의 존재 여부를 판단할 수 있으며, 데이터 전처리 과정에서의 평가 지표로써 활용할 수 있다.

이처럼 베이지안 신경망은 모델의 불확실성을 정량화할 수 있다는 점에서 타 기계학습 알고리즘과 차별화된 강점을 보유하고 있으며, 이를 이용하여 기계학습 및 건물에너지 시뮬레이션에 관한 다양한 파생 연구가 진행될 수 있을 것으로 판단된다.

참 고 문 헌(16pt)

1 Afram, A., Janabi-Sharifi, F. (2014). Review of modeling methods for HVAC systems. Applied Thernal Engineering. 67, pp.507-519.

2 Afram, A., Janabi-Sharifi, F., Fung, A. S., Raahemifar, K. (2017). Artificial neural network based model predictive control and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system. Energy and Buildings. 141, pp.96-113.

3 ASHRAE (2014). ASHRAE Guideline 14-2014: measurement of energy and demand savings. American Society of Heating, Refrigerating and Air-conditioning Engineers, Atlanta, GA

4 Barber, D., & Bishop, C. M. (1998). Ensemble learning in Bayesian neural networks. Nato ASI Series F Computer and Systems Sciences, 168, (pp. 215-238).

5 Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight uncertainty in neural networks. arXiv preprint arXiv:1505.05424.

6 Der Kiureghian, A., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter?. Structural safety, 31(2), 105-112.

7 Every, P. M. V., Rodriguez, M., Jones, C. B., Mammoli, A. A., Martínez-Ramón, M. (2017). Adavanced detection of HVAC faults using unsupervised SVM novelty detection and Gaussian process models. Energy and Buildings. 149, pp.216-224.

8 Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In international conference on machine learning (pp. 1050-1059).

9 Gal, Y. (2016). Uncertainty in deep learning. PhD Thesis. University of Cambridge.

10 Hinton, G. E., & Van Camp, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. In

Proceedings of the sixth annual conference on Computational learning theory (pp. 5-13).

11 Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.

12 Kendall, A. G., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision?. In Advances in neural information processing systems (pp. 5574-5584).

13 Kendall, A. G. (2019). Geometry and uncertainty in deep learning for computer vision. PhD Thesis. University of Cambridge.

14 Kim, H., Jung, D. C., Choi, B. W. (2019). Exploiting the Vulnerability of Deep Learning-Based Artificial Intelligence Models in Medical Imaging: Adversarial Attacks. Journal of the Korean Society of Radiology, 80 (2):259

15 Kwon, Y., Won, J. H., Kim, B. J., & Paik, M. C. (2018). Uncertainty quantification using bayesian neural networks in classification: Application to ischemic stroke lesion segmentation.

16 Kwon, Y., Won, J. H., Kim, B. J., & Paik, M. C. (2020). Uncertainty quantification using bayesian neural networks in classification: Application to biomedical image segmentation. Computational Statistics & Data Analysis, 142, 106816.

17 Neal, R. M. (1995). Bayesian learning for neural networks. PhD thesis, University of Toronto.

18 Nikolaidou, E., Wright, J., & Hopfe, C. J. (2015, December). Early and detailed design stage modelling using Passivhaus design; what is the difference in predicted building performance. In BS2015, 14th Conference of International Building Performance Simulation Association, Hyderabad, India, December 7 (Vol. 9, pp. 2166-2173).

19 Park, S. H., Ahn, K. U., Hwang, S. H., Choi, S. K., Park, C. S. (2019).

Machine learning vs. hybrid machine learning model for optimal operation of a chiller. Science and Technology for the Built Environment. 25, pp.209-220.

20 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

21 Tax, D. M. J. and Duin, R. P. W. (2004). Support vector data description. Machine Learning, 54(1), (pp. 45–66).

22 Williams, C. K. (1997). Computing with infinite networks. NIPS.

23 Williams, C. K., & Rasmussen, C. E. (2006). Gaussian processes for machine learning (Vol. 2). Cambridge, MA: MIT press.

Abstract

Epistemic and Aleatoric Uncertainty of Bayesian Neural Network Model for a Chiller

Kim Jae Min

Department of Architecture and Architecture Engineering

The Graduate School

Seoul National University

Because the machine learning model is a black-box model, it is difficult to quantify the causality between inputs and outputs. In addition, the model is influenced by its inherent uncertainty in describing a system's behavior of interest. In order for a machine learning model to be reliable, its prediction performance as well as uncertainty must be quantified together.

The uncertainty of the machine learning model is divided into epistemic uncertainty and aleatoric uncertainty. Epistemic uncertainty is caused by lack of data or knowledge. In contrasts, aleatoric uncertainty is uncertainty caused by intrinsic randomness of natural phenomena such as sensing noises or malfunction of the system.

Most machine learning algorithms used in computer science do not offer the information about model confidence or uncertainty. Bayesian

Neural Network (BNN) is a useful tool to describe stochastic characteristics of deep learning models by estimating distributions of the models' weights. However, BNN is still unpractical because of their computational costs. In this thesis, Yarin Gal(2016)'s methodology is used to obtain uncertainty of BNN model in more practical way, which uses Monte Carlo estimation with dropout neural networks.

In this thesis, the BNN models were developed for a compression chiller in an existing office building with BEMS data, and then epistemic and aleatoric uncertainties were analyzed. It is found that both uncertainties are significant in the simulation model even though the model's accuracy is satisfactory with the CVRMSE of less than 30%. It is suggested that before attempting to apply the machine learning model to real applications, the both uncertainties must be carefully analyzed. It is recommended that the both uncertainties can be reduced by adding more data as well as removing outliers.

keywords : Building energy simulation, Machine learning, Bayesian Neural Network, Epistemic Uncertainty, Aleatoric uncertainty, Support Vector Data Description

Student Number : 2018-28645